Vol. 43, Parts 3 and 4 December 1956

UN. OF MIC N
DEC 21 1956
SCIENCE LIBRARY
PR
Science
QH
301
.B615
v.43
pts.3/4

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITOR

M. G. KENDALL

in consultation with

HARALD CRAMÉR
F. N. DAVID
R. C. GEARY
J. B. S. HALDANE
H. O. HARTLEY
D. G. KENDALL

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued 4 December* 1956]

A volume of *Biometrika* containing about 500 pages, with plates and tables, will be published annually in two half-yearly issues.

Papers for publication should be sent either to

PROFESSOR E. S. PEARSON

Department of Statistics, University College, London, W.C. 1

or if more convenient to

PROFESSOR M. G. KENDALL

London School of Economics, Houghton Street, London, W.C. 2

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each. Order forms for separates are sent to authors with proofs of their papers.

The Subscription price, payable in advance, is:

Inland: 45*s.* net per volume Abroad: 54*s.* net per volume

including packing and postage

BACK ISSUES

Volumes 20–43

These may be obtained from the BIOMETRIKA OFFICE, at the following prices:

Vols. 20–38: £5 per volume Vols. 39–43: £3. 3*s.* per volume

packing and postage 2s. per volume

Bound volumes: £1 extra per volume Binding cases: 10*s.* each

Cheques should be made payable to *Biometrika*, crossed "a/c BIOMETRIKA TRUST" and sent to

THE SECRETARY
BIOMETRIKA OFFICE
UNIVERSITY COLLEGE
GOWER STREET, LONDON, W.C. 1

to whom all orders for series, single copies and offprints should be addressed.

All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

Volumes 1–19

Permission for reprinting has been granted to Messrs Wm. Dawson & Sons Ltd. Volumes 1–13 are now ready for distribution and Volumes 14–19 are in course of preparation. Would librarians and others wishing to have copies please place their orders now with:

WM. DAWSON & SONS LTD.,
4 DUKE STREET, MANCHESTER SQUARE,
LONDON, W. 1

VOLUME 43, PARTS 3 AND 4 DECEMBER 1956

STUDIES IN THE HISTORY OF PROBABILITY AND STATISTICS

III. A NOTE ON THE HISTORY OF THE GRAPHICAL PRESENTATION OF DATA

BY ERICA ROYSTON

Division of Research Techniques, London School of Economics and Political Science

1. The cartoon of the business man despondently watching his production curve disappearing off the chart on the wall and down through the floor into the office below appears to most people to be a joke that has worn slightly thin. Nevertheless, the very fact that graphical representation can become the subject of a cartoon shows how completely it has come into everyday usage. The graph is now generally accepted as one of the clearest and most effective ways of presenting data, whether for consumption by the layman or the specialist. Many volumes have been published on the art of graphical presentation, on the appropriate method to be used for any specific purpose and on the advantages and disadvantages of various approaches to the subject. It seems strange therefore that so little curiosity has been displayed regarding the historical origins of the graphical representation of data.* In fact, these historical origins are not at all clear.

2. As in most other historical studies of this nature it is virtually impossible to state categorically that a given method was introduced by any one person or at any one moment of time. The most that one may hope to do is to show that such a method was being used at a certain time and to investigate the author's claims, if any, to originality. Apart from this, some of the underlying concepts can be traced back to their possible origins.

3. For the purpose of this note the graphical representation of data will be taken to mean the geometric and graphical presentation of factual data as distinct from the graphical plotting of mathematical functions. Although the development of the technique of using Cartesian co-ordinates with one axis representing time seems to date back only a few centuries, the spatial representation of time originated much earlier. There exist instances of the movement of stars, or, more precisely, the inclinations of the planetary orbits being plotted as a function of time, as early as the tenth century (Funkhouser, 1936), and the emergence of written music represents one of the earliest instances of the use of time-series.†

4. The basic idea of using co-ordinates to determine the location of a point in space dates back to the Greeks at least, although it was not until the time of Descartes that mathematicians systematically developed the idea. For the purpose of tracing the origins of graphical representation the emergence of the concept of plotting mathematical functions

* While holding the part-time appointment of Professor of Geometry at Gresham College (1890–4) Karl Pearson gave a series of twelve lectures under the heading 'The Geometry of Statistics'. The synopsis which survives refers to Playfair as the father of the subject but the lectures themselves appear to have been lost. See E. S. Pearson (1938, p. 142.)

† Neumes (*νευματα*), the basis of present-day musical notes, were at first only approximately measured in terms of time. The duration of notes was not fixed precisely until the twelfth and thirteenth centuries, when the so-called Franconian reform (after Franco of Cologne's *Ars Cantus Mensurabilis*) took place. Together with the gradual introduction of the bar and a fixed beat in the bar, music became what may be termed a true time-series, where notes were graphically presented, using the tone as ordinate against the time as abscissa.

of the type $y=f(x)$ is important. The most usual form of time-series is, in fact, little more than this, except that usually some form of frequency is plotted as an observed, not necessarily strictly mathematical, function of time. Cartesian co-ordinates are therefore one of the foundations on which the modern graph of data is based. It is interesting to note in passing that William Playfair (1801) justifies plotting money against time as follows:

> This method has struck several persons as being fallacious because geometrical measurement has not any relation to money or to time yet here it is made to represent both. The most familiar and simple answer to this objection is that if the money received by a single man in trade were all guineas and every evening he made a single pile of all the guineas received during the day, its height would be proportioned to the receipts of that day, so that by this plain operation time, proportion and amount would be physically combined.

As will be shown, however, this picturesque description is merely an explanation to the layman, for Playfair himself almost certainly approached the subject as an extension of the use of rectangular co-ordinates, which by his time were familiar in mathematics.

5. One of the early users of graphical representation in statistics was A. F. W. Crome. He was born in Germany in 1753, the third of twenty children. His father was a clergyman and, following in his footsteps, Crome studied theology. During his studies he acted as tutor to the children of General von Holzendorf and later to those of Karl Alexander von Bismarck. He passed his examinations in 1771 and in 1783 became lecturer in geography and history at Dessau. In 1783 he became a tutor to the 16-year-old Prince of Dessau, and finally in 1786 he was made ordinary Professor of Statistics and Public Finance at Giessen University, where he remained till he retired in 1831. He died two years later.

6. With an academic background such as this it is hardly surprising to learn that Crome first evolved his system of geometric representation as an aid to teaching. Crome was, in modern terminology, an economic geographer rather than a statistician. His works are descriptive rather than analytical. Although the *Allgemeine Deutsche Biografie* calls him a 'pioneer in statistics', it is probably using the word in its original meaning, i.e. as data referring to states. Thus the adjective 'statistical', which appears in nearly all the titles of his works, e.g. *Geographisch-statistische Darstellung der Staatskräfte*, merely seems to indicate that some numerical data are given. Thus, for example, in *Über die Grösse und Bevölkerung der sämtlichen Europäischen Staaten* (1785), Crome gives a detailed description of the geographical 'vital data' of most European states, chiefly population figures and areas. To make the importance of such data clearer he devised his *Grössen Karte* which was of much the same type as that illustrated in Fig. 1. (The latter is taken from a similar treatise dealing with the German states.) He justifies the use of this geometrical representation as follows (1785):

> If therefore one does not wish to limit one's knowledge of geography to knowing the names of the cities, provinces, rivers and monarchs, but also wishes to include an overall view of the condition and might of the various empires and states in order to get a good grasp of their might, size and culture, and to compare these among themselves, one must get the largest possible overall view of the area and population of the European states—this is the purpose of the attached chart.
>
> On the one hand the eye can take in some idea of the relative size of the European states by just looking at a general map of this part of the world. But on the other hand the outline of the states is such that they cannot easily be compared.
>
> ...The proportions of the different sizes can however be more easily seen and grasped if they are brought before the eye in the form of a drawing, because the imagination is thus stimulated, than if these merely appeared in the form of numbers, especially when these consist of many digits as is often the case with areas of states....

7. The method Crome employed in these charts was simple. He expressed the area of each of the states with which he was dealing as a square proportional to that area. He then drew these squares inside each other in such a manner as to keep the vertical sides of the squares, which are of course proportional to the square root of the area represented, in scale. When the area of a state was relatively much larger than the others he broke the scale in much the same manner as is done on modern graphs.

8. The idea of representing data in this way seems to have been thought completely new, for Crome went into great detail in describing the uses and misuses of his charts. For example, he warns his readers that there is no justification for assuming that, because the squares were drawn one inside the other, the outside square, or rather the country represented by it, was larger than all those within it added together. It appears that Crome was led to the idea by the comparison of geographical areas on maps; in fact, he used spatial representation of magnitude, not the linear magnitudes moving through time of the Cartesian approach.

There is, however, one example of graphical representation being used by one of Crome's contemporaries and compatriots, namely, G. R. von Gothe's *Höhen Tableau* (Altitude Table) (1813). This is an imaginary panoramic scene, complete with trees and animals, where most of the highest known mountains of Europe and America appear in proportion, their altitudes being given by reference to a scale on the vertical axis.

9. In 1817 Crome wrote:

> I have used this method with great success in my 35 year career as Docent. To this end I published in 1782 my first *Produkten Karte* of Europe, which was followed in 1785 by the *Grössen Karte*. This was issued in an improved form in 1792 as the *Verhältniss Karte* of Europe.

It seems that the *Produkten Karte* referred to differed little from the later ones,* and it is a reissue of the *Verhältniss Karte* that is illustrated here as Fig. 1. It appears as part of a much larger chart which also gives lists of data referring to military strength, etc.

10. On the base of this chart, published in 1820, is a diagram that Crome seems to have used here for the first time. An extract of it is shown as Fig. 2. It consists basically of a series of circles each representing the density of population in the state to which it refers, the density being inversely proportionate to the area of the circle. The various half-tangents and extended radii denote total population and national income, the lines being drawn in different colours to distinguish them from one another. Thus total population is denoted as follows:

The tangent on the right-hand side of several circles indicates the population in millions as measured on the scale, the radii leaning to the right indicate 100 thousands, and the thin black lines also leaning to the right thousands. (N.B. The length of these slanting lines is misleading, as the quantity is indicated by the number of sections of the vertical scale crossed.) These quantities are then added together to give the total population. In fact, each of these lines represents a group of digits from the original figure.

Thus, for example, Luxemburg (third from left) has a population density of 1981 per German square mile (from the text below the circle). Its population is indicated by the

* This chart is missing in the copies of Crome's works available both in the British Museum and the British Library of Political Science.

extended radius sloping to the right which shows 2·5 units, the units in this case being 100,000 (see above). Total population is thus 250,000. The national income is shown as follows:

The tangent on the left indicates 1 unit, i.e. 1 million Gulden, the extended radius sloping to the left indicates 5 units, i.e. $5 \times 100{,}000$ Gulden. Total national income is thus $1{,}000{,}000 + 500{,}000 = 1{,}500{,}000$ Gulden. National income per head, as indicated by the radius extended downwards is 6 Gulden.

11. Although a little involved and misleading and on the whole giving only a very poor overall picture, this method is perhaps more precise than if the population and national income figures had all been plotted to scale. Basically this could be taken as a rather crude attempt at a bar chart, the circles being merely a generalization of the squares Crome used earlier.

12. Crome nowhere explicitly claims that he invented the methods he used, although implicitly he does so several times, as, for example, in the last passage quoted above. Nevertheless, this last method of circles, although it could merely have been a generalization of the squares, seem similar enough to that used by Playfair (1801) a few years earlier (Fig. 3) to warrant closer examination of the possibility that Crome did get some of his ideas from the latter.

13. William Playfair was born in 1759 near Dundee, and was also the son of a clergyman. His father died when William was 13, his brother John, who later became famous as a result of his work in the field of mathematics and geology, taking charge of the family. William Playfair was apprenticed to Andrew Meikle, inventor of the threshing machine. In 1780 he became a draughtsman for Boulton and Watt, and while with them he patented two inventions, the 'Eldorado sash' and a machine for making fretwork on silver tea-trays. Backed by these inventions Playfair opened a shop in London, which soon failed. In 1789 he succeeded Joel Barlow as agent of the Scioto (Ohio) Land Company, a rather unsavoury concern which sold shares in land which may or may not have existed. While in Paris Playfair probably helped to capture the Bastille; he is known also to have saved Duval from the mob in 1791. In the next year allegations concerning his 'mismanagement' of the Ohio Company, coupled with some rather plain speaking against the Revolutionaries, forced Playfair to leave Paris for Frankfurt.

14. Some of Playfair's doings appear a little odd. For example, in 1793, while in Frankfurt, a French emigré gave him details of a system of signalling by means of the semaphore system. He quickly made models of the apparatus—a large clock-like dial, several of which were to be erected across the country to form a chain whereby messages could be relayed—and sent them to the Duke of York. Thereafter he claimed to have introduced the system to England. In actual fact, R. L. Edgeworth (1797) had already used it in 1767 to relay the results of a race at Newmarket to his home some miles away. On his return to London, Playfair opened a 'Security Bank' to facilitate small loans by subdividing large securities. This very soon collapsed, and from 1795 Playfair lived only by writing. Among other things he claimed to have warned the government of Napoleon's intention to escape from Elba. After Waterloo he returned to Paris to edit *Galigni's Messenger*. While there he got involved in a case of libel, and to escape the sentence of three months' imprisonment he fled to London. The remaining years before his death in 1823 he spent in London writing pamphlets.

15. There is nothing of the studious or academic background of Crome here, and it is therefore rather surprising that what Playfair wrote was good sound economics and, moreover, that his works were illustrated with some extremely good graphs, histograms and pie diagrams. He wrote mainly works on general descriptive economics, one of his favourite subjects being the international balance of trade. Most of the graphs he used were plotted time-series, chiefly export and import figures expressed in millions of pounds; but he also used circles to present spatial magnitudes after the manner of Crome. His earliest work containing such graphs seems to be his *Commercial and Political Atlas*, published in 1786. This consists of graphs of exports and imports represented by coloured lines, the gap between them being termed the balance in favour of England and usually also coloured. He produced such graphs for trade to and from most of Britain's foreign markets. In the same book there are also some very good bar diagrams showing the amount of exports to and imports from each country plotted against time.

16. Playfair's professed aim in publishing his charts was to make statistics, presumably again in the sense of data relating to state-craft, a little more palatable. Statistics, one may gather from the following extract from Playfair's *Statistical Breviary* (1801), were not thought of very highly.

> . . . for no study is less alluring or more dry and tedious than statistics, unless the mind and imagination are set to work or that the person studying is particularly interested in the subject; which is seldom the case with young men in any rank in life.

17. Most of Playfair's later diagrams were merely improvements of his earlier work. Fig. 4 shows one such graph where two series have been superimposed upon one another, one as a graph and the other as a histogram (the two series here being weekly wages and the price of a quarter of wheat (1821)). He illustrated his *British Family Antiquity* with several beautifully executed bar diagrams, which differed from the true frequency graphs used elsewhere by the fact that only the presence or absence of a given factor is plotted against time. This type of diagram, Playfair conceded, had long been used in chronology.

18. Fig. 3 shows an extract from an earlier example of Playfair's work. Here the areas of European states are expressed as circles and in some cases subdivided as pie diagrams. The line on the left-hand side of the circles indicates population in millions and that on the right-hand side national revenue in millions of pounds. The slope of the line joining these two is intended to indicate the approximate ratio of one to the other. Playfair nowhere states this and is presumably aware of the fact that the differing diameters of the circles make any exact comparison impossible. This particular figure has been mentioned earlier as closely resembling Crome's work (Fig. 2) and therefore warranting the examination of the possibility of there being some connexion between the two writers. Certainly this was possible, for Crome and Playfair were contemporaries and, moreover, both wrote in the same field. Playfair's diagram was, however, published before Crome's, and it seems unlikely that, having had sufficient flair to produce perfect time-series graphs, he should have copied Crome. Conversely, however, it seems equally unlikely, even were it possible, that Crome should have taken a basically clear and simple diagram of Playfair's and made such a complicated diagram as Fig. 2 out of it—unless, of course, he was working under the impression that he was improving on Playfair's work. On the face of it, with no direct evidence either way, it seems that if there was any connexion between the two it was Crome who was copying from Playfair and not vice versa. But it is at least equally possible that the two worked entirely independently.

19. There is nothing crude or clumsy about Playfair's work and it seems surprising that one man should have developed it to such a high pitch—for, unlike Crome, Playfair makes definite claims to originality, calling himself the 'inventor of linear arithmetic' as he termed graphical representation. In 1796, for instance, he writes:

> I confess I was long anxious to find out whether I was actually the first who applied the principles of geometry to matters of finance as it has long been applied to chronology with great success. I am now satisfied, upon due enquiry, that I was the first, for during the 11 years I have never been able to learn that anything of a familiar nature had ever before been thought of...

and later (1805):

> The impression is not only simple, but it is as lasting in retaining as it is easy in receiving. Such are the advantages claimed for the invention 20 years ago, when it first appeared. The claim has been allowed and not objected to so far as the inventor knows, either in this or in any other country.

And as an explanation of how he came to think of such graphical representation (1805);

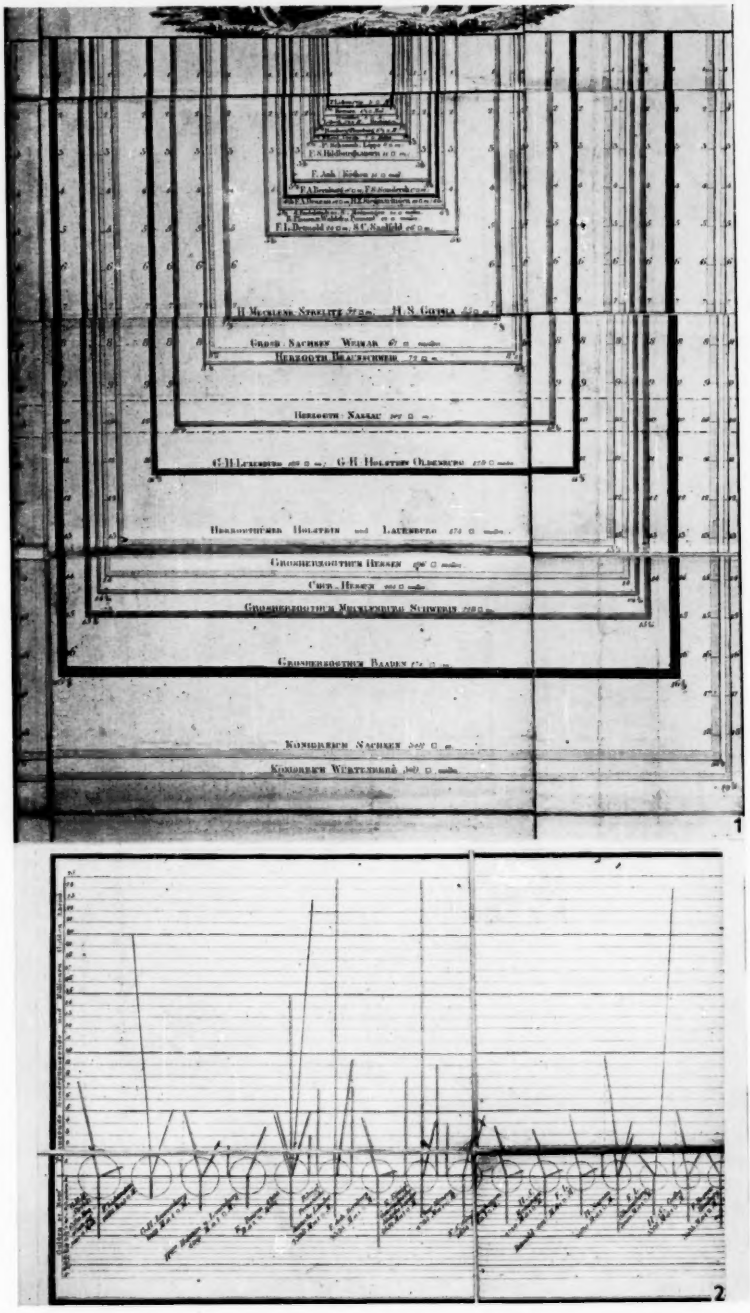
> I think it well to embrace this opportunity, the best I have had, and perhaps the last I ever shall have, of making some return (as far as acknowledgement is a return) for an obligation, of a nature to be repaid by acknowledging publicly, that to the best and most affectionate of brothers I owe the invention of these charts.
>
> At a very early period of my life, my brother, who in a most exemplary manner maintained and educated the family his father left, made me keep a register of a thermometer, expressing the variation by lines on a divided scale. He taught me to know that whatever can be expressed in numbers may be represented by lines. The chart of the thermometer was on the same principle with those given here, the application only is different. The brother to whom I owe this now fills the Natural Philosophical Chair at the University of Edinburgh.

20. Were it not for the somewhat doubtful background revealed by his biographers, including among other matters the bogus claim concerning the semaphore, this sounds honest enough. Certainly there did not seem to be any counterclaims made by his contemporaries, for Playfair's work was acclaimed both in this country and in France, where even the Academy of Sciences 'testified its approbation of this application of geometry to accounts' (Playfair, 1798).

21. Apart from the influence exerted on Playfair by his brother, there is, however, another clue to be found in his biography. In 1780 Playfair worked as a draughtsman to Boulton and Watt. Now Watt is well known as the first engineer who illustrated in diagrammatic form the work done in a steam-engine cylinder by graphing pressure against volume, and one of his draughtsmen would undoubtedly have come into contact with graphical methods of presenting many kinds of motion. It seems very possible that Playfair's real originality lay, not in the kind of diagram he used, but in its application to descriptive statistics and economics.

22. On the whole, therefore, it seems quite possible that William Playfair, if not the 'inventor' of graphical representation as we know it to-day, was the first to introduce it into statistics; and that Crome independently had much the same ideas so far as concerns spatial presentation, but did not carry them out quite as well. The only reasons for suspecting Playfair's authenticity are his occasional dubious activities in other fields, and in the circumstances it is perhaps advisable to accord him the benefit of the doubt.

23. Little that Playfair did would seem to have been beyond the capabilities or scope of men like Graunt, Petty or Halley a century earlier. It seems somewhat surprising that the



Figs. 1 and 2. Extracts from Crome's *Geographisch-statistische Darstellung*, 1820.

(*Facing p.* 246)

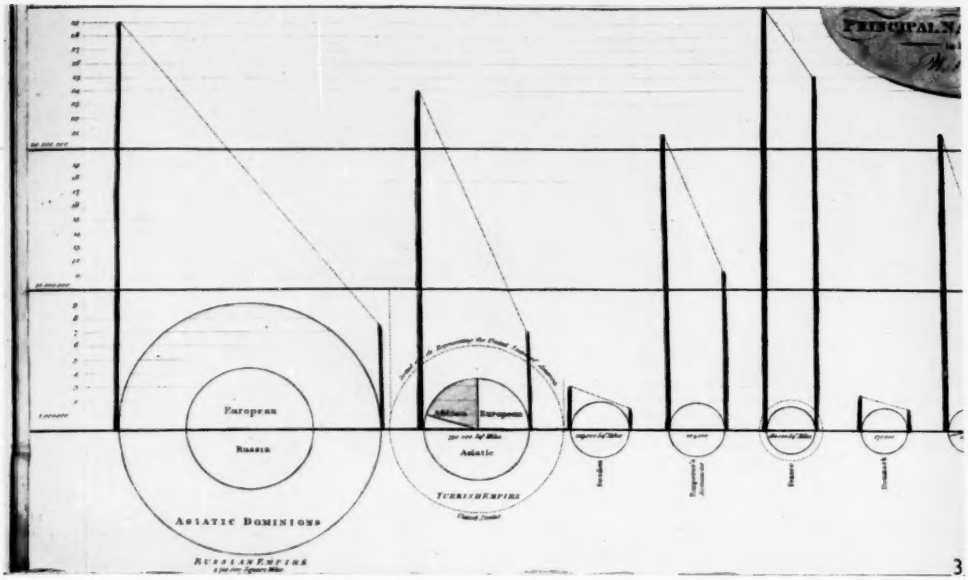


Fig. 3. Extract from Playfair's *Statistical Breviary*, 1801.

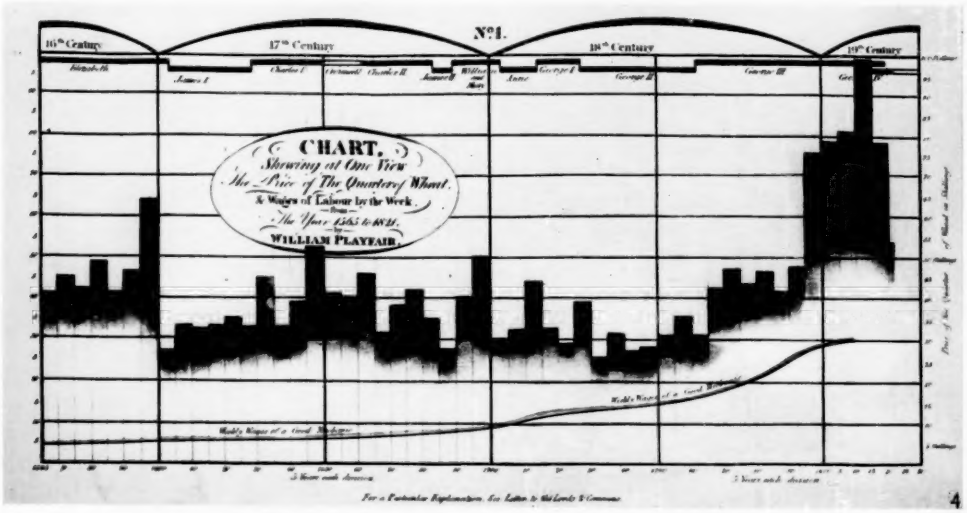


Fig. 4. Extract from Playfair's *A Letter on our Agricultural Distress*, 1821.

graphical presentation of statistics should have had to wait until the end of the eighteenth century. The explanation may well be that until the middle of that century there was very little to present. To quote Playfair himself (1801):

> Statistical knowledge, though in some degree searched after in the most early ages of the world, has not till within these last 50 years become a regular object of study.

REFERENCES

CROME, A. F. W. (1785). *Über die Grösse und Bevölkerung der sämtlichen Europäischen Staaten.* Leipzig.

CROME, A. F. W. (1820). *Geographisch-statistische Darstellung der Staatskräfte.* Leipzig.

EDGEWORTH, R. L. (1797). *A Letter to the Rt. Hon. Earl of Charlemont on the Tellograph.* Dublin.

FUNKHOUSER, H. G. (1936). A note on a 10th century graph. *Osiris*, **1**. Bruges.

PEARSON, E. S. (1938). *Karl Pearson.* Cambridge University Press.

PLAYFAIR, W. (1786). *The Commercial and Political Atlas.* London.

PLAYFAIR, W. (1796). *For the Use of the Enemies of England.* London.

PLAYFAIR, W. (1798). *Linear Arithmetic.* London.

PLAYFAIR, W. (1801). *Statistical Breviary.* London.

PLAYFAIR, W. (1805). *An Enquiry into the Permanent Causes of the Decline and Fall of Powerful Nations.* London.

PLAYFAIR, W. (1808). *British Family Antiquity.* London.

PLAYFAIR, W. (1821). *A Letter on our Agricultural Distress.* London.

VON GOTHE, G. R. (1813). Höhen Tableau. *Allgemeine Geographische Ephemeriden*, **41**. Weimar.

STUDIES IN THE HISTORY OF PROBABILITY AND STATISTICS

IV. A NOTE ON AN EARLY STATISTICAL STUDY OF LITERARY STYLE

By C. B. WILLIAMS, Sc.D., F.R.S.

In *Biometrika* for January 1939, G. Udny Yule discussed the frequency distribution of sentence length in samples of the writings of different authors. After showing that each author had a fairly characteristic distribution, he turned to the value of the method in cases of uncertain or disputed authorship. Thus, in the case of *De Imitationi Christi*, he showed that the frequency distribution of sentences with different numbers of words more closely resembled that of works by Thomas à Kempis than that of works by de Gerson.

In *Biometrika* for March 1940 I showed that the skew distribution found by Yule could be brought almost to a symmetrical form by using a geometric or logarithmic scale for the number of words per sentence, thereby simplifying the mathematical comparisons. In this note I mentioned that some years previously (about 1935) I had made a number of frequency distributions from different authors using the number of letters per word as the variable, but that I had not found any striking differences. I considered Yule's use of the number of words per sentence as a better technique, giving a greater range of possible variation and comparison.

In a letter written to me in June 1939 Yule said: 'I booked up some ten years ago a number of distributions of word-length by the number of syllables only. Monosyllables are always considerably in the majority (if I remember rightly I omitted "a" and "the"), and different authors diverged a good deal, but, so far as I can recall, the range from Bunyan to a *Times* Leader was not so very striking.'

Neither Yule nor myself was aware that quite extensive investigations in this line had been made and published in summary more than fifty years previously, giving frequency distributions of word lengths (by the number of letters per word) for several authors, and suggesting that similar distributions of the numbers of syllables per word, or the number of words per sentence, might well help to throw light on cases of doubtful or disputed authorship.

Through the kindness of Mr Rushworth Fogg of Glasgow I was put on the track of a paper published in 1901 by Thomas Corwin Mendenhall,* in which he gives a reference to a still earlier paper published in 1887, both of which I have been able to examine.

Mendenhall states in his first paper (1887) that five or six years previously he had seen a suggestion in a book by Augustus de Morgan, possibly his *Budget of Paradoxes*, that it might be possible to identify the author of a book, a poem or a play by the average length of the words used in the construction. Mendenhall, however, considered that the method which he had adopted in this publication, of using the frequency distribution of words of different lengths, was better, as while the average number of letters per word is easily obtainable from his data, the shape of the distribution provides considerably increased possibilities of comparison.

Augustus de Morgan was Professor of Mathematics at University College, London. His *Budget of Paradoxes* was first printed as weekly notes in the *Athenaeum* and republished in

* See biographical note on p. 255 below.

book form in 1872, after the author's death. I have examined the second edition (1915), but, although it contains a few references to cases of disputed authorship, I cannot find any suggestion about the use of the average number of letters per word. It may be in one of his earlier works, or perhaps in one of his *Athenaeum* notes that was not reprinted in book form.

It is interesting to note than Mendenhall, who was primarily a physicist, was attracted to the frequency distribution technique by its resemblance to spectroscopic analysis, which in 1887 was much to the fore in scientific circles. He writes: 'It is proposed to analyse a composition by forming what may be called a "word spectrum" or "characteristic curve" which shall be a graphic representation of the arrangement of words according to their length and the relative frequency of their occurrence.' The mathematics of the comparison of frequency distributions was very little understood at the time when he was writing.

Mendenhall first discusses samples taken from different books by the same author to see if they resemble each other sufficiently closely to make comparisons between one author and another likely to be profitable. Most of the evidence is given in the form of thirteen diagrams and, unfortunately, in only very few cases are actual numbers presented.

Mendenhall's first seven graphs deal with various combinations of ten samples, each of one thousand words, from Dickens's *Oliver Twist* and Thackeray's *Vanity Fair*. In his second figure the distribution of five separate samples of 1000 words from *Oliver Twist* are shown superimposed, and there is no doubt as to their general resemblance. In another (Fig. 1 in this paper) he shows one graph for the whole 10,000 words from *Oliver Twist* and another for *Vanity Fair*. There is very little difference in the average length of the words (Dickens, 4·324; Thackeray, 4·481), but *Vanity Fair* has rather more words of 3 and of 7–10 letters, while *Oliver Twist* has more of 1, 2 and 4–6 letters. Mendenhall was somewhat disappointed by the lack of difference and commented 'it is certainly suprising that...so close an agreement should be found. This is particularly striking in the words of 11, 12 and 13 letters, the numerical composition of which is as follows:

Number of letters	11	12	13
Dickens	85	57	29
Thackeray	85	58	29'

Undaunted by this small difference he next tried two groups of words from John Stuart Mill's *Political Economy* and his *Essay on Liberty*, in which he 'expected to find more longer words than in the novelists'. 'But I confess to considerable surprise in finding from the very beginning that, although on the whole the anticipation was realized, the word which occurred most frequently was not the three-letter word, as with both Dickens and Thackeray, but the word of two letters.' The explanation he says 'is to be found in the liberal use of prepositions in sentence-building'. The results, given in two separate diagrams, are here combined into one (Fig. 2).

Mendenhall next studied two addresses given by a Mr Edward Atkinson on 'Labour Questions' to two different audiences, one consisting of working men and the other students of a Theological College. There was 'a marked difference in style', but the word-length distributions (from two samples of 5000 words) were very similar. He comments that 'Mr Atkinson's composition was remarkable in the shortness of the words used'. The average length was 4·298 letters; which is, however, only 0·044 shorter than the samples from Dickens.

For comparison with all the above studies of works in the English language, Mendenhall gave a distribution of the first 5500 words in Caesar's *Commentaries*, in Latin. He finds a mean word length of 6·065 letters and an entirely different form of curve with peaks at 2, 5 and 7 letters (see Fig. 3). This is of course connected with the Latin construction of adding to the main root for inflexions instead of using additional small words.

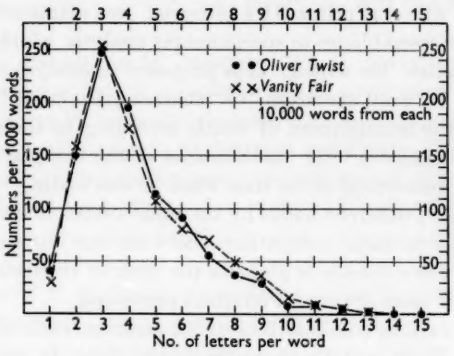


Fig. 1

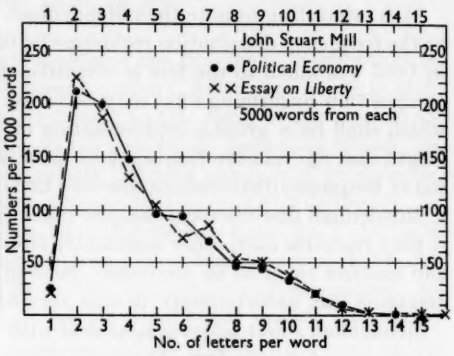


Fig. 2

Fig. 1. Samples of 10,000 words each from Dickens's *Oliver Twist* and Thackeray's *Vanity Fair*. Redrawn from Mendenhall (1887, fig. 7).

Fig. 2. Samples of 5000 words each from two works of John Stuart Mill. Redrawn from Mendenhall (1887, figs. 8, 9).

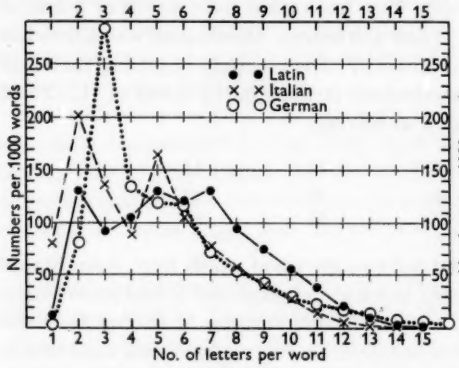


Fig. 3

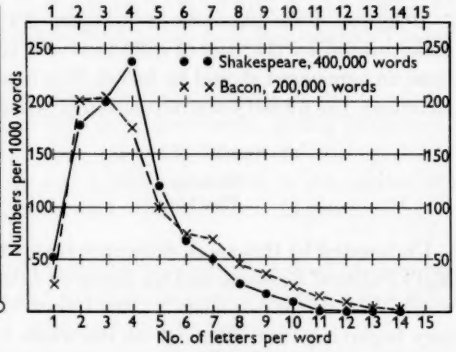


Fig. 4

Fig. 3. Examples of distribution of word length in languages other than English. (1) Latin, (2) Italian, (3) German. Redrawn from Mendenhall (1887, fig. 13 and 1901, fig. 2).

Fig. 4. Comparison of frequency distribution of word length in two very large samples from the plays of Shakespeare and works of Bacon. Redrawn from Mendenhall (1901, fig. 7).

He considers that, for a really reliable estimate of the characteristic curve for an author, a sample curve of 100,000 words might be necessary, and concludes his first paper as follows:

'Many interesting applications of the process will suggest themselves to almost every reader; the most notable, of course, being the attempt to solve questions of disputed authorship, such as exist in reference to the letters of Junius, the plays of Shakespeare, and other less widely known examples. It might also be used in comparative language studies,

in tracing the growth of a language, and in studying the growth of the vocabulary from childhood to manhood.'

'If striking differences are found between the curve of known and suspected compositions of any writer, the evidence against identity of authorship would be quite conclusive. If the two compositions should produce curves which are practically identical, the proof of a common origin would be less convincing for it is possible though not probable, that two writers might show identical curves.'

It was not until 14 years later than Mendenhall returned to the problem in a paper in *Popular Science Monthly*, published in December 1901. In this he repeats some of the discussion and diagrams from his earlier paper dealing with Dickens, Thackeray, John Stuart Mill and Mr Atkinson; but in addition to his earlier diagrams showing analysis of a Latin work he gives examples of single authors in Italian, Spanish, French and German (see Fig. 3). The French and Spanish curves, possibly by the idiosyncrasies of the authors chosen, have their peaks in the 2-letter words; the Italian has two peaks at 2 and 5 letters; while the German has a peak at 3 letters, but with more longer words, reaching one word of 27 letters.

After this introduction, he settles down to a discussion of the value of his technique in the study of the authorship of the plays of Shakespeare. For this the length of nearly two million words were counted from the works of Shakespeare and of some of his contemporaries. Most unfortunately the data are all condensed into half a dozen small diagrams, and not one table of the actual numbers is given. Does his evidence still exist, hidden away somewhere?

Mendenhall says: 'The result from the start, with the first group of 1000 words, was a decided surprise. Two things appeared from the beginning: Shakespeare's vocabulary consisted of words whose average length was a trifle below four letters, less than any writer of English previously studied; and his word of greatest frequency was the four-letter word, a thing never before met with' (Fig. 4).

A comparison of the diagrams with those of Thackeray and Dickens shows that Shakespeare had a higher proportion of words with 1, 2, 4 and 5 letters, and a lower proportion of words with 3 letters and of 6 letters upwards; which accounts for the fact that while his peak is higher, his average number of letters per word is lower. In modern terminology, he would have a smaller standard deviation.

Altogether about 400,000 words were counted including 'in whole or in part, nearly all his most famous plays', and it was found that this characteristic curve is most persistent—that based on first 50,000 words differing very little from the whole count. In a diagram giving two examples of 200,000 words each, it is practically impossible to separate the two lines, in spite of the fact that Mendenhall says the differences have been of necessity slightly exaggerated in order to make them show at all!

A comparison was next made of Shakespeare's prose and his poetry as exemplified by *The Rape of Lucrece* and *Venus and Adonis*. The prose gave more shorter words, particularly of 2 letters, and fewer words of 5, 6 and 7 letters; but both gave the characteristic peak at the 4-letter word. Mendenhall writes: 'At first this was thought to be a general characteristic of his time, but this was found not to be so.'

A study was then made of a number of works by Francis Bacon, including his *Henry VII* and his *Advancement of Learning*, with a total of nearly 200,000 words. The frequency distribution was quite different from that of Shakespeare (see Fig. 4) with the peak at the 3-letter word, with more 2-letter words, fewer with 4, 5 and 6 letters, and more longer words with 7–13 letters. Mendenhall here comments that 'the reader is at liberty to draw any

conclusion he pleases from this diagram. Should he conclude that, in view of the extraordinary difference in these lines, it is clear that Bacon could not have written the things ordinarily attributed to Shakespeare... the question still remains, who did?'

An examination of the works of Ben Johnson, in two groups of 75,000 words, showed once more a peak at the 3-letter word; but an extensive study of the plays of Beaumont and Fletcher showed that on the final average the number of 4-letter words was slightly greater than those of 3 letters, although the excess was by no means persistent in smaller samples. The final curve was not unlike that of Shakespeare, and Mendenhall suggested that the 'lack of persistency of form among small groups' might be due to the dual authorship.

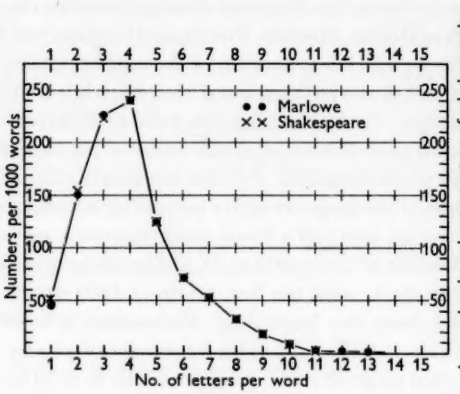


Fig. 5

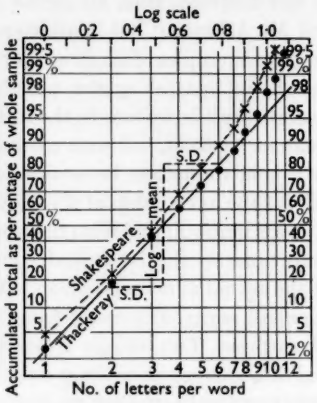


Fig. 6

Fig. 5. Comparison of frequency distribution of word length in two large samples from the plays of Shakespeare and of Christopher Marlowe. Redrawn from Mendenhall (1901, fig. 9).

Fig. 6. The number of letters per word, on a logarithmic scale, from works of Thackeray and Shakespeare (as shown by Mendenhall) plotted against the accumulated total as a percentage of the whole sample on a probability scale. It indicates some resemblance to a log-normal distribution for words up to about 8 letters, but differing above this level.

When, however, he turned his attention to the plays of Christopher Marlowe 'something akin to a sensation was produced among those engaged in the work'. 'In the characteristic curve of his plays Marlowe agrees with Shakespeare about as well as Shakespeare agrees with himself' (Fig. 5).

Finally, Mendenhall pointed out that a dramatic composition *Armada Days* written by Prof. Shaler of Harvard, in which the author endeavoured to compose in the spirit and the style of the 'Elizabethan days', gave a curve (from only about 20,000 words) with 'excess of the 4-letter word and in other respects decidedly Shakespearian'. Mendenhall does not give this curve or any figures.

Discussion

We are not concerned here so much with the results that Mendenhall obtained, or with their repercussions, but rather with the general value of the technique. There appears to be little doubt that he was the first to act on the suggestion of de Morgan, and that his own method of using the frequency distribution, instead of merely the average length of word, was a distinct improvement, although the average length would not normally be given

to-day without the standard deviation. The skew form of the curve makes this latter measure less reliable than it might otherwise be.

Mendenhall's sampling method was to take blocks of 1000 words each 'at the beginning of the volume and, after a few thousand words had been counted, the book was opened near the middle and the count continued'. This method is not above reproach, but, in view of the large number of samples and the general close resemblances, it is unlikely that a more randomized method would produce any measurably different result. In the case of the plays of Shakespeare the sampling was large enough to justify the statement that it included nearly all the most famous plays 'in whole or in part'.

That Mendenhall appreciated the difference between the statistical method and evidence based on selected phraseology believed to be characteristic is clear from the following quotations: 'The chief merit of the method consisted in the fact that its application required no exercise of judgement' and that 'characteristics might be revealed which the author could make no attempt to conceal, being himself unaware of their existence'; and again, 'the conclusions reached through its use would be independent of personal bias, the work of one person in the study of an author being at once comparable with the work of any other'.

That Mendenhall saw the wide range of possibilities is clear from his statement: 'it is hardly necessary to say that the method is not necessarily confined to the analysis of a composition by means of word-length: it may equally be applied to the study of syllables, of words in sentences, and in various ways.' And I have already quoted his suggestion as to its value in comparative studies.

Two additional comments may be of interest. The curve of the frequency distribution of words of different lengths is in every case skew, with the peak usually at 3 or 4 letters per word, and the tail running off generally to 15 or 16, but sometimes to higher than this. In my contribution to the study already mentioned, I showed that by the use of a logarithmic scale the skew distributions of sentence length became approximately symmetrical, and so the distribution resembled a log-normal. It is of interest to see if Mendenhall's figures for word length show a similar relation. We can, however, note beforehand that the length of a sentence is under the conscious control of a writer, who may stop when he pleases. The lengths of words are not so controlled and selection of words for reason of their length alone is not likely to occur.

Most unfortunately only three sets of numbers are given by Mendenhall, all in his first 1887 paper. They are for 1000 words in *Oliver Twist* and two sets of the same size for *Vanity Fair*. Taking the latter we find that the accumulated totals up to each successive number of letters per word, expressed as percentages of the whole, are as shown in Table 1.

When these results are plotted on to log-probability paper the result is as shown in Fig. 6. There is an approximately straight-line relation up to about 8 letters per word, but above that there is a definite departure. The straight-line portion suggests a log-normal distribution with a mean log at 0·53 and a standard deviation of approximately 0·26. On an arithmetic scale this is equivalent to a geometric mean of about 3·4 and a standard deviation of $\times$ or $\div$ 1·8.* The arithmetic mean is 4·5 letters per word.

* When a frequency distribution is skew on an arithmetic grouping of the data but approximately symmetrical when a geometric scale is used, the standard deviation cannot be expressed on an arithmetic scale as '+ or −'. The use of the expression '3·4, $\times$ or $\div$ by 1·8' implies that approximately 33% of the observations will be between 3·4 and $3{\cdot}4 \times 1{\cdot}8$; 33% will be between 3·4 and $3{\cdot}4 \div 1{\cdot}8$; and approximately 17% will be above, and below, these limits.

In his second paper Mendenhall gives five graphs showing Shakespeare's frequency distribution per 1000 words in comparison with other authors. With a lens and a fine scale it is possible to read the numbers to about three units, but unfortunately the results so obtained from the five diagrams do not agree. This is possibly because (as he admits in one case) Mendenhall exaggerated the differences in the diagrams in order to separate the two lines. I have made an estimate from each of the five diagrams, and the average values are given in Table 2.

Table 1

No. of letters per word	No. of words out of 2000	Accumulated no. of words total	Accumulated total % of 2000
1	58	58	2·9
2	315	373	18·7
3	480	853	42·7
4	351	1204	60·2
5	244	1448	72·4
6	154	1602	80·1
7	152	1754	87·7
8	100	1854	92·7
9	63	1911	98·9
10	43	1960	98·0
11	16	1976	98·8
12	15	1991	99·6
13	4	1995	99·8
14	5	2000	100·0

Table 2

Letters	Words	Letters	Words	Letters	Words
1	47·6	6	71·2	11	3·4
2	175·8	7	52·6	12	2·0
3	225·0	8	31·6	13	1·0
4	237·6	9	18·4	14	0·4
5	124·4	10	9·0	—	—

When the accumulated totals are plotted on log-probability paper as in the previous case, the result (Fig. 6) indicates a fairly regular departure from the straight line, although once again the break is more distinct above 7 letters per word.

I have also attempted to get, from Mendenhall's diagram of five samples of 1000 words each from *Oliver Twist*, some measure of the error of his results.

The frequency distribution of words of certain lengths in the five samples is approximately as given in Table 3.

If the size of the sample were increased sixteen times—to 80,000 words—without altering the pattern of the material sampled, the S.E. of the mean would be reduced to a quarter of the above—or approximately 2·1, 1·1, 2·1, 1·5 and 0·8 for words of 3–6 letters respectively. The error is smaller in the less frequent words, but greater in proportion to the mean.

Thus for the comparison of two samples of this size—assuming the same order of variation in each—the S.E. of the difference would be approximately 1·4 times the above or 2·9, 1·7, 2·1 and 1·1. Thus differences in number of words per 1000 would have to be of the order of 5·7, 3·3, 4·1 and 2·1 to be significant at the 1 in 20 level, and 7·5, 4·4, 5·4 and 2·8 to be significant at the 1 in 100 level. The five samples on which the above rough estimate is made were, however, consecutive samples of 1000 words from one work; when different works, written at different periods by the same author, are combined the error of the mean would almost certainly be greater.

A careful examination of Mendenhall's diagram giving the comparison of the distributions of Shakespeare and Marlowe suggests measurable differences only in words up to 5 letters, Marlowe differing from Shakespeare approximately as follows: 1 letter, 5 less; 2 letters, 3 less; 3 letters, 3 more; 4 letters, no difference; and 5 letters, 5 more. All the other

Table 3

No. of letters	Five samples	Mean	S.E. of mean for 5000 words
3	221, 232, 236, 254, 268	242·2	8·36
4	170, 175, 183, 186, 198	182·4	4·82
5	95, 102, 120, 122, 123	112·4	5·80
6	83, 92, 94, 97, 103	93·8	3·28

word lengths are indistinguishable in the diagram. These differences may have been exaggerated in the diagram. The numbers are in words per 1000 in large samples—in the case of Shakespeare over 200,000 words, but in the case of Marlowe the size of samples is not given. On the other hand, the comparison of Bacon with Shakespeare (see Fig. 5) shows a difference of nearly 60 words per thousand with 4-letter words.

It would seem likely that real differences between authors would show themselves as sequences of departures in the same direction for several consecutive word lengths. One can imagine one author differing from another in an unconscious preference for longer words or for shorter words, but it is unlikely that one author would prefer words of, say, 11 letters in preference to 12, while another author would prefer the 12 to the 11. Thus, rapid changes of departure directions in sequence would be less convincing than blocks of departures of similar sign as evidence of real differences, and would be more likely to be due to error.

Mendenhall, in his 1887 paper, calls attention to the fact that in an analysis of Dickens's *Christmas Carol*, words of 7 letters appeared to be unduly numerous, due to the fact that the character 'Scrooge', frequently referred to, is a word of this length. It would be desirable to leave names of persons and places out of any tabulation.

BIOGRAPHICAL NOTE

Thomas Corwin Mendenhall was born in Ohio on 4 October 1841 and died there on 22 March 1924. He was the descendant of a Benjamin Mendenhall who emigrated from England (probably from Wiltshire) in 1686 to join Penn's Colony and who settled at Concord,

Pennsylvania. T. C. Mendenhall spent some early years as a school teacher, but in 1873 became the first Professor of Physics and Mechanics at the newly founded Ohio Agricultural and Mechanical College. He was Professor of Physics in the University of Tokyo from 1878 to 1881 and in the Ohio State University from 1881 to 1886. He then became President of the Rosa Polytechnic Institute in Indiana and was elected the following year to the National Academy of Science. After a few years as Superintendent of the U.S.A. Coast and Geodetic Survey he became President of the Worcester Polytechnic Institute, where he remained until his retirement at the age of 60.

His biography by Henry Crow (*Biograph. Mem. Nat. Acad. Sci., Wash.*, **16**, 331–51), to which I am indebted for the above information, lists about sixty publications in physics, particularly geophysics, units of electrical measurement, state boundary lines in the U.S.A. and many other related subjects. The first of the two papers at present under discussion is listed, but not the second. There is no mention of his interest in the statistics of literary style, but it is said that he left in MSS. about 900 pages of an autobiography which has never been published. If sti" [illegible] it might repay study.

EFERENCES

Mendenhall, T. C. (1887). The characteristic curves of composition. *Science*, **9** (214, supplement), 237–49.

Mendenhall, T. C. (1901). A mechanical solution of a literary problem. *Pop. Sci. Mon.* **9**, 97–105

De Morgan, A. (1872). *A Budget of Paradoxes*. London (2nd edition 1915).

Williams, C. B. (1940). A note on the statistical analyses of sentence length as a criterion of literary style. *Biometrika*, **31**, 356–61.

Williams, C. B. (1952). Statistics as an aid to literary studies. *Penguin Science News*, no. 24, pp. 99–106.

Yule, G. U. (1939). On sentence-length as a statistical characteristic of style in prose. *Biometrika*, **30**, 363–84.

A GOODNESS OF FIT TEST FOR SPECTRAL DISTRIBUTION FUNCTIONS OF STATIONARY TIME SERIES WITH NORMAL RESIDUALS

BY A. M. WALKER

Design and Analysis of Scientific Experiment, 6 Keble Road, Oxford

1. INTRODUCTION

In recent years problems of statistical inference associated with the spectral analysis of stationary time series have been studied by a number of authors (see, for example, Bartlett, 1950, 1954; Grenander, 1951; Grenander & Rosenblatt, 1953; Whittle, 1951, 1954). In most of these problems the aim has been to obtain information about the spectral density function of the series, which is usually assumed to be absolutely continuous. Grenander & Rosenblatt (1953) give asymptotic confidence bands for the spectral distribution function, but still require the assumption of an absolutely continuous spectral density function.

The present paper adopts an approach which is associated directly with the spectral distribution function, and yields a general large sample goodness of fit test of the hypothesis that a stationary Normal series has a specified spectral distribution function which is not necessarily absolutely continuous but may have certain points of discontinuity.

2. THE BASIC TEST STATISTICS AND THEIR MAIN PROPERTIES

Let $\{X(t)\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ be a time series which is stationary to the second order. We may assume without loss of generality that $X(t)$ is measured from its mean, i.e. that $E[X(t)] = 0$.

Then it is well known that we can write

$$X(t) = \int_{-\pi}^{\pi} e^{it\omega} dZ(\omega) \tag{1}$$

(the integral being interpreted in the mean square sense), where $Z(\omega)$ is a complex-valued orthogonal process defined over $(-\pi, \pi)$ with†

$$\left.\begin{aligned} &E[|\, Z(\omega_2) - Z(\omega_1)\,|^2] = F(\omega_2) - F(\omega_1), \\ &E[\{Z(\omega_4) - Z(\omega_3)\}\{Z(\omega_2) - Z(\omega_1)\}^*] = 0, \\ &\text{when the intervals } (\omega_1, \omega_2),\ (\omega_3, \omega_4) \text{ do not overlap,} \end{aligned}\right\} \tag{2}$$

$F(\omega)$ being the spectral distribution function (not necessarily normalized to unity). $F(\omega)$ is non-negative and non-decreasing over $(-\pi, \pi)$ and is related to the autocovariance function $\sigma_X^2 \rho(\tau) = E[X(t)\, X(t+\tau)]$ by the equation

$$\sigma_X^2 \rho(\tau) = \int_{-\pi}^{\pi} e^{i\omega\tau}\, dF(\omega). \tag{3}$$

† The symbol * denotes complex conjugate.

The inversion formula corresponding to (1) is, with $Z(0) = 0$,

$$Z(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} [(1-e^{-i\omega r})/(ir)]\, X(r), \tag{4}$$

for every continuity point ω of F and therefore of Z. Here the infinite sum is to be interpreted as a mean-square limit $\lim_{n\to\infty} \sum_{r=-n}^{n}$ and $(1-e^{-i\omega r})/(ir)$ is to be replaced by ω when $r = 0$. This suffices to determine $Z(\omega)$ uniquely for all ω in $(-\pi, \pi)$, since at points of discontinuity it can always be defined to be continuous on the right (corresponding to $F(\omega)$ being continuous on the right).

Let $A(\omega)$ and $B(\omega)$ be the real and imaginary parts of $Z(\omega)$. Then we have

$$E[A(\omega_2) - A(\omega_1)]^2 = E[B(\omega_2) - B(\omega_1)]^2 = \tfrac{1}{2}[F(\omega_2) - F(\omega_1)], \tag{5}$$

$$\left.\begin{aligned} E[\{A(\omega_4) - A(\omega_3)\}\{A(\omega_2) - A(\omega_1)\}] = E[\{B(\omega_4) - B(\omega_3)\}\{B(\omega_2) - B(\omega_1)\}] = 0 \\ \text{when } (\omega_1, \omega_2),\ (\omega_3, \omega_4) \text{ do not overlap,} \end{aligned}\right\}, \tag{6}$$

and

$$\left.\begin{aligned} E[\{A(\omega_4) - A(\omega_3)\}\{B(\omega_2) - B(\omega_1)\}] = 0 \\ \text{for any intervals } (\omega_1, \omega_2),\ (\omega_3, \omega_4). \end{aligned}\right\} \tag{7}$$

These results do not follow directly from (2), but can easily be established using (4) (cf. also Doob, 1952, p. 482).

Suppose now that we are given a set of observations at $2n+1$ consecutive times, say $X(-n)$, $X(-n+1)$, ..., $X(0)$, $X(1)$, ..., $X(n)$. Then we can approximate to $A(\omega)$ and $B(\omega)$ by the finite sums

$$A_{2n}(\omega) = \frac{1}{2\pi} \sum_{r=-n}^{n} \left(\frac{\sin \omega r}{r}\right) X(r), \tag{8}$$

and

$$B_{2n}(\omega) = \frac{1}{2\pi} \sum_{r=-n}^{n} \left(\frac{1-\cos \omega r}{r}\right) X(r). \tag{9}$$

If n is fairly large, we should expect that (5), (6) and (7) will remain approximately true when $A_{2n}(\omega)$ and $B_{2n}(\omega)$ are substituted for $A(\omega)$ and $B(\omega)$, provided that ω is a continuity point of F. This suggests that quantities of the form $A_{2n}(\omega_2) - A_{2n}(\omega_1)$ or $B_{2n}(\omega_2) - B_{2n}(\omega_1)$ may be used to construct tests of the goodness of fit of the observations to the spectral distribution function $F(\omega)$.

We therefore take such quantities as our basic test statistics. Since B_{2n} is an even function of ω and A_{2n} an odd function we need consider only non-negative values of ω; in view of this it is convenient to express our results in terms of the spectral distribution function $F_+(\omega)$ defined over $(0, \pi)$, which is such that

$$d_\omega F_+(\omega) = 2d_{-\omega} F(-\omega) = 2d_\omega F(\omega) \quad (0 < \omega < \pi) \quad \text{and} \quad \sigma_X^2 \rho(\tau) = \int_0^\pi \cos \omega\tau\, dF_+(\omega). \tag{10}$$

The main properties of these statistics, corresponding to equations (5), (6) and (7), are as follows:

(i) Let (ω_1, ω_2) be a continuity interval for $F_+(\omega)$, i.e. an interval whose end-points are continuity points of $F_+(\omega)$, with $\omega_1 \gg \pi/n$ and $\pi - \omega_2 \gg \pi/n$, and let n be so large that

$$F_+(\omega_i + \lambda\pi/n) - F_+(\omega_i - \lambda\pi/n) \quad (i = 1, 2), \quad \text{where} \quad \lambda = O(1),$$

is small compared with $F_+(\omega_2)-F_+(\omega_1)$. Then to a good approximation

$$E[A_{2n}(\omega_2)-A_{2n}(\omega_1)]^2 = E[B_{2n}(\omega_2)-B_{2n}(\omega_1)]^2 = \tfrac{1}{4}[F_+(\omega_2)-F_+(\omega_1)]. \tag{11}$$

Also if $\omega_2 \gg \pi/n$ and $F_+(\omega_2+\lambda\pi/n)-F_+(\omega_2-\lambda\pi/n)$, $F_+(\lambda\pi/n)$ are small compared with $F_+(\omega_2)$, we may put $\omega_1=0$ in (11) provided that F_+ has no discontinuity at $\omega=0$. A similar result holds with $\omega_2=\pi$, provided that F_+ has no discontinuity at $\omega=\pi$.

(ii) Let (ω_1,ω_2) and (ω_3,ω_4) be continuity intervals for $F_+(\omega)$, and *either* $\omega_3-\omega_2 \gg \pi/n$ or $\omega_3-\omega_2 = O(\pi/n)$ and $F(\frac{1}{2}(\omega_2+\omega_3)+\lambda\pi/n)-F(\frac{1}{2}(\omega_2+\omega_3)-\lambda\pi/n)$, with $\lambda=O(1)$, be small compared with $F(\omega_2)-F(\omega_1)$, $F(\omega_4)-F(\omega_3)$. Then to a good approximation

$$\left.\begin{aligned} E[\{A_{2n}(\omega_4)-A_{2n}(\omega_3)\}\{A_{2n}(\omega_2)-A_{2n}(\omega_1)\}] &= 0,\\ E[\{B_{2n}(\omega_4)-B_{2n}(\omega_3)\}\{B_{2n}(\omega_2)-B_{2n}(\omega_1)\}] &= 0.\end{aligned}\right\} \tag{12}$$

(iii) For any intervals (ω_1,ω_2), (ω_3,ω_4)

$$E[\{A_{2n}(\omega_4)-A_{2n}(\omega_3)\}\{B_{2n}(\omega_2)-B_{2n}(\omega_1)\}] = 0. \tag{13}$$

This last result, which is exact, follows at once from the fact that in

$$\sum_{r,s=-n}^{n}\left(\frac{\cos r\omega_3-\cos r\omega_4}{r}\right)\left(\frac{\sin s\omega_2-\sin s\omega_1}{s}\right)\sigma_X^2\rho(r-s),$$

the terms with $(r,s)=(r_0,s_0)$ and $(r,s)=(-r_0,s_0)$ cancel, there being no contribution from $(r,s)=(0,0)$. The derivation of (i) and (ii) is more complicated, and certain difficulties arise in assessing the accuracy of the approximations (it will be noted that the conditions under which this may be expected to be high have not been stated at all precisely). The discussion of these at this point would mean a fairly lengthy digression from the main argument of the paper, and is therefore postponed until the concluding section (§ 6).

Now let $\{X(t)\}$ be a Normal series. Then $A_{2n}(\omega)$, $B_{2n}(\omega)$ are Normal random variables, and hence, from (11) and (13),

$$[A_{2n}(\omega_2)-A_{2n}(\omega_1)]^2+[B_{2n}(\omega_2)-B_{2n}(\omega_1)]^2$$

is, under the conditions stated in (i) above, distributed approximately as

$$\tfrac{1}{4}\{F_+(\omega_2)-F_+(\omega_1)\}\chi^2_{(2)},$$

$\chi^2_{(m)}$ denoting a random variable distributed in the standard χ^2 form with m degrees of freedom. Also, using (12), it follows that if $0=\omega_0<\omega_1<\omega_2\ldots<\omega_k=\pi$ is a subdivision of the interval $(0,\pi)$ such that $\omega_0,\omega_1,\ldots,\omega_k$ are continuity points of $F_+(\omega)$, and the 'power' (increment in F_+) in each interval (ω_i,ω_{i+1}) is large compared with that in $(\omega_i-\lambda\pi/n,\omega_i+\lambda\pi/n)$ and $(\omega_{i+1}-\lambda\pi/n,\omega_{i+1}+\lambda\pi/n)$, $\lambda=0(1)$ $(i=0,1,\ldots,k-1)$, then

$$[A_{2n}(\omega_{i+1})-A_{2n}(\omega_i)]^2+[B_{2n}(\omega_{i+1})-B_{2n}(\omega_i)]^2 \quad (i=0,1,\ldots,k-1) \tag{14}$$

are approximately distributed as $\frac{1}{4}\{F_+(\omega_{i+1})-F_+(\omega_i)\}\chi^2_{(2)}$ independently of one another.†

It should be noted that

(i) if $E[X(t)]$ is not specified *a priori* we may replace $X(r)$ in (8) and (9) by $X(r)-\overline{X}$, where $\overline{X}$ is the mean of the set of observations. This introduces an additional term in (8) which is of order $n^{-\frac{1}{2}}$, and hence for sufficiently large n does not affect the approximations (11) and (12);

† When 0 or π are not continuity points of $F_+(\omega)$ we take $\omega_0>0$ or $\omega_k<\pi$.

(ii) the assumption that the number of observations is odd is not important since with an even number, $2n+2$ say, we can, in defining $A_{2n}(\omega)$ and $B_{2n}(\omega)$, either omit one of the end observations, the loss of information being asymptotically negligible, or make the limits of summation $-n$, $n+1$, which gives an additional term of order n^{-1}.

It does not seem possible to make any general statement about the distributions of $A_{2n}(\omega)$ and $B_{2n}(\omega)$, and hence those of the statistics (14), without the assumption that $\{X(t)\}$ is Normal. We might, for example, wish to consider the case of the general discrete linear process, defined by $X(t) = \sum_{u=-\infty}^{\infty} g(t-u)\,\epsilon(u)$, the $\epsilon(u)$ being mutually independent with a common distribution, but the limiting distributions of $A_{2n}(\omega)$, $B_{2n}(\omega)$ as $n \to \infty$ will clearly depend on the form of this distribution, e.g. for the independent process $\{X(t)\} = \{\epsilon(t)\}$, the mth cumulant of $A_{2n}(\omega) \to \left\{\omega^m + 2\sum_{r=1}^{\infty}\left(\frac{\sin \omega r}{r}\right)^m\right\}\kappa_m(\epsilon)$.

3. The goodness of fit criteria

Let the null hypothesis be that $\{X(t)\}$ has a specified spectral distribution function $F_+(\omega)$. Then we can construct statistics

$$T_i = 4[\{A_{2n}(\omega_{i+1}) - A_{2n}(\omega_i)\}^2 + \{B_{2n}(\omega_{i+1}) - B_{2n}(\omega_i)\}^2]/\{F_+(\omega_{i+1}) - F_+(\omega_i)\} \quad (i = 0, 1, \ldots, k-1), \tag{15}$$

which for sufficiently large n will, to a good approximation, be distributed independently as multiples of $\chi^2_{(2)}$, these multiples being unity on the null hypothesis, and either greater or less than unity on the alternative hypotheses. Equivalently $s_i^2 = \frac{1}{2}T_i$ will be approximately independent variance estimates each with $f_i = 2$ degrees of freedom, the corresponding variances σ_i^2 being equal to unity on the null hypothesis and having arbitrary values on the alternative hypotheses. From this property we can easily derive suitable criteria for testing goodness of fit.

The most obvious one to use is the standard likelihood ratio criterion

$$\begin{aligned} M = -2[L - L_{\max.}] &= \sum_{i=0}^{k-1} f_i(s_i^2 - 1 - \log s_i^2) \\ &= \sum_{i=0}^{k-1} (T_i - 2\log T_i) + 2k(\log 2 - 1). \end{aligned} \tag{16}$$

On the null hypothesis we have $M \sim (\frac{7}{6})\,\chi^2_{(k)}$, using the approximation due to Bartlett (1937) (which should be sufficiently accurate except perhaps for small k, when it may be advisable to use the better approximation, due to Box (1949), based on the standard F distribution). For large k it is worth while modifying the χ^2 approximation to $M \sim 1{\cdot}118\chi^2_{(1{\cdot}032k)}$, which, if we neglect the effect of the approximations of § 2, reproduces the mean and variance of M exactly, $\kappa_3(M)$ with an error of about 3 %, and $\kappa_4(M)$ with an error of about 8 % (this is easily verified from the expression $k\{\log\Gamma(1-2\phi) - 2\phi - (1-2\phi)\log(1-2\phi)\}$ for the cumulant generating function of M, $\log E(\exp M\phi)$).

An alternative criterion is obtained by assessing the significance of each T_i separately, using a two-tailed test, and then combining the results by Fisher's method (see, for example, Fisher, 1941, p. 95). Thus if

$$p_i = \exp -\tfrac{1}{2}T_i, \quad \text{we let} \quad y_i = \begin{cases} 1-p_i & (p_i \geqslant \frac{1}{2}) \\ p_i & (p_i < \frac{1}{2}) \end{cases},$$

and refer $U = \sum_{i=0}^{k-1} u_i$, where $u_i = -2\log 2y_i$, to a χ^2 distribution with $2k$ degrees of freedom, high values being considered significant. With this procedure there is the advantage that from the individual values of the p_i, the intervals (ω_i, ω_{i+1}) for which there are highly significant departures of the numerators of (15) from their expectations on the null hypothesis are immediately identifiable. However, the form of the power function of the test may be unsatisfactory. This is indicated by its behaviour when k is large (but still of course $\ll n$). For it can be shown, using the Normal approximation to the distribution of U, that for alternative hypotheses with $\epsilon_i = (1/\sigma_i^2) - 1$ of order $k^{-\frac{1}{2}}$, the power function is then approximately equal to

$$1 - \Phi(\xi_\alpha + 0{\cdot}42\theta), \tag{17}$$

where $\Phi(x) = \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy/\sqrt{(2\pi)}$, ξ_α is such that $\Phi(\xi_\alpha) = 1-\alpha$, α being the significance level of the test, and $\theta = k^{-\frac{1}{2}} \sum_{i=0}^{k-1} \epsilon_i$ (see Appendix, § 1). Clearly the test is ineffective for such hypotheses with $\theta > 0$, since then $(17) < \alpha$.

This difficulty is overcome by modifying the U test as follows. Let $U_1 = \Sigma_1 u_i$, $U_2 = \Sigma_2 u_i$, where Σ_1 denotes summation over the k_1 values of i for which $p_i \geqslant \frac{1}{2}$, and Σ_2 summation over the remaining $k_2 = k - k_1$ values for which $p_i < \frac{1}{2}$, and refer U_1, U_2 to χ^2 distributions with $2k_1$ and $2k_2$ degrees of freedom respectively, rejecting the null hypothesis if either is significantly large. This modified criterion is, for large k, sensitive to deviations ϵ_i of order $k^{-\frac{1}{2}}$ giving positive or negative values of θ (Appendix, § 2 (9)). The comparison of the power functions of the modified U test and the M (likelihood ratio) test is in general a difficult problem, but the fact that, for large k, M is sensitive to deviations ϵ_i of order $k^{-\frac{1}{4}}$ (Appendix, § 3 (12)), supports the view that the modified U test is, on the whole, the more powerful.

There are of course other criteria which are more effective for particular types of alternative hypothesis. Among these may be mentioned $\max T_i$, the largest of the T_i, which would be appropriate for alternative hypotheses under which a large jump in the spectral distribution function (or a large peak in the spectral density function, its derivative) occurred at some point. The distribution of $\max T_i$ under the null hypothesis is obviously given by

$$\Pr\{\max T_i > x\} = 1 - (1 - e^{-\frac{1}{2}x})^k.$$

These tests can also be applied when, as is usually the case, $F_+(\omega)$ is not completely specified under the null hypothesis, but contains a number of parameters which have to be estimated from the data, provided that consistent estimators of these parameters are available. For then the effect of the substitution of the estimators for the parameters occurring in the denominators of the T_i is negligible when n is sufficiently large. For example, the variance of the series, σ_X^2, will seldom be given *a priori*, at most the normalized spectral distribution function $F_+(\omega)/F_+(\pi)$ being specified, and we can usually substitute the estimator

$$\hat{\sigma}_X^2 = \sum_{r=-n}^{n} (X(r) - \bar{X})^2/(2n),$$

which is certainly consistent, having a mean-square error of order n^{-1}, when $\Sigma_r |\rho(\tau)|$ converges. Again, if on the null hypothesis $\{X(t)\}$ is a linear autoregressive process, defined by the stationary solution of

$$X(t) + a_1 X(t-1) + \ldots + a_p X(t-p) = Y(t),$$

where the $Y(t)$ are independent Normal variables with zero mean and constant variance, and the roots of $z^p + a_1 z^{p-1} + \ldots + a_p = 0$ have moduli less than unity, we can substitute the usual least squares estimates of the parameters $a_1 \ldots a_p$.

With the likelihood ratio test we can, in fact, avoid the estimation of σ_X^2 by taking the null hypothesis to be the equality of the σ_i^2, their common value being unspecified. M then becomes Bartlett's criterion for testing the homogeneity of a set of variance estimates. Again the test of the significance of the largest T_i can be made independent of σ_X^2 by taking the criterion to be $g = \max T_i \Big/ \sum_{i=0}^{k-1} T_i$, whose distribution on the null hypothesis, given by

$$\Pr(g > x) = \sum_{r=1}^{[1/x]} (-1)^{r-1} \binom{k}{r} (1-rx)^{r-1},$$

$[1/x]$ denoting the integral part of x, was first obtained by Fisher (1929).

It is worth noting that if the spectral density function $f_+(\omega)$ exists and is continuous, then when the difference between its upper and lower bounds in (ω_i, ω_{i+1}) is sufficiently small, we may, in calculating T_i, replace $F_+(\omega_{i+1}) - F_+(\omega_i)$ by $(\omega_{i+1} - \omega_i)\,\frac{1}{2}(f_+(\omega_{i+1}) + f_+(\omega_i))$.† The advantage of doing so lies in the fact that the process is then usually of the generalized autoregressive form for which

$$X(t) + a_1 X(t-1) + \ldots + a_p X(t-p) = Y(t) + b_1 Y(t-1) + \ldots + b_q Y(t-q),$$

giving immediately

$$f_+(\omega) = (\sigma_Y^2/\pi) \,|\, (1 + b_1 e^{i\omega} + \ldots + b_q e^{qi\omega}) / (1 + a_1 e^{i\omega} + \ldots + a_p e^{pi\omega}) \,|^2;$$

although, since $f_+(\omega)$ is a rational function of $\cos\omega$, $F_+(\omega)$ can be obtained explicitly in terms of elementary functions, it may be quite a complicated expression.

4. Comparison with tests based on periodogram intensities

The periodogram intensity for angular frequency ω is given by

$$\left.\begin{aligned} I_{2n+1}(\omega) &= H_{2n+1}(\omega)\, H^*_{2n+1}(\omega), \\ \text{where} \qquad H_{2n+1}(\omega) &= \sqrt{\frac{2}{2n+1}} \sum_{r=-n}^{n} e^{i\omega r} X(r). \end{aligned}\right\} \tag{18}$$

It has been shown that when the spectral density $f_+(\omega)$ exists and is continuous near ω_i $(i = 1, 2, \ldots, k-1)$, the statistics $I_{2n+1}(\omega_i)/\pi f_+(\omega_i)$ have the same distributional properties as the T_i, i.e. for large n are, to a good approximation, distributed as $\chi^2_{(2)}$, provided that $\omega_{i+1} - \omega_i \gg \pi/n$ and some suitable condition is imposed on the autocorrelations $\rho(\tau)$—the convergence of $\Sigma_\tau \,|\, \rho(\tau) \,|$ is certainly sufficient (see, for example, Bartlett, 1950, p. 4). It might be thought that there would be a direct connexion between T_i and $I_{2n+1}(\omega_i)$, but this is not so since

$$\{A_{2n}(\omega_2) - A_{2n}(\omega_1)\}^2 + \{B_{2n}(\omega_2) - B_{2n}(\omega_1)\}^2 = 2\pi^2(2n+1) \left| \int_{\omega_1}^{\omega_2} H_{2n+1}(\omega)\, d\omega \right|^2.$$

† Our requirement on the spacing of the ω_i for the results (11) and (12) to hold to a good approximation then becomes $\omega_{i+1} - \omega_i \gg \pi/n$.

Goodness of fit tests using these statistics may thus be obtained in the same way as those of § 3. Such tests have been proposed by several authors; for example, the likelihood ratio test is discussed by Bartlett (1950, pp. 7–9), and the Fisher g test by Whittle (1952a, p. 47) and Sargan (1953). Their application need not be confined to the case of null hypotheses for which the spectral distribution function is absolutely continuous. For we can write $F_+(\omega) = F_+^{(1)}(\omega) + F_+^{(2)}(\omega)$, where $F_+^{(1)}(\omega)$ is absolutely continuous and $F_+^{(2)}(\omega)$ is a step function with a finite or enumerably infinite set of discontinuity points (cf. Bartlett, 1955, p. 163), and hence can always choose the ω_i so that $f_+(\omega_i)$ exists (in practice this will mean that $f_+(\omega_i)$ is also continuous near ω_i).

When $f_+(\omega)$ is a sufficiently smooth function the power of one of these periodogram intensity tests should be approximately equal to that of the corresponding test based on the T_i with intervals (ω_i, ω_{i+1}) such that the periodogram frequencies are $\frac{1}{2}(\omega_i + \omega_{i+1})$. For let the spectral density be $f_+^{(0)}(\omega)$ on the null hypothesis and $f_+^{(0)}(\omega) + f_+^{(1)}(\omega)$ on an alternative hypothesis. Then when the variation of $f^{(0)}, f^{(1)}$ with ω is approximately linear in (ω_i, ω_{i+1}), the expectations of the variance estimates s_i^2 on the alternative hypothesis become approximately equal to $1 + \left[f_+^{(1)}\left(\frac{\omega_i + \omega_{i+1}}{2}\right) \Big/ f_+^{(0)}\left(\frac{\omega_i + \omega_{i+1}}{2}\right)\right]$ in both cases, and these determine the power functions. The tests of § 3 should be better for detecting large peaks in $f_+(\omega)$ of unknown location. For with the intensity tests, such peaks will have appreciable probabilities of being detected only when they occur at frequencies whose differences from some ω_i are of order n^{-1}.

However, when $f_+^{(0)}(\omega)$ is continuous for all ω, intensity tests of greatly increased power may be obtained by using frequencies $\omega_j = 2\pi j/(2n+1)$ $(j = 1, 2, \ldots, n)$ (cf. Whittle, 1952a,b; Sargan, 1953). For then the intensities $I_{2n+1}(\omega_j)$ and $I_{2n+1}(\omega_{j'})$ $(j' \neq j)$ are asymptotically uncorrelated, at least when (as will usually be the case) we can write $X(t) = \sum_{u=0}^{\infty} g(u)\,\epsilon(t-u)$, where the $\epsilon(t)$ are independent Normal variables with zero mean and constant variance σ_ϵ^2, and $g(u) \to 0$ exponentially as $u \to \infty$; this follows most easily from the asymptotic relation

$$I_{(X)}(\omega) \sim (\pi f_+(\omega)/\sigma_\epsilon^2)\, I_{(\epsilon)}(\omega) \tag{19}$$

between the intensities for $\{X(t)\}$ and $\{\epsilon(t)\}$ (see Bartlett, 1955, p. 279). It is reasonable to expect that (as assumed by Whittle and Sargan) the asymptotic distributions of the criteria will not be affected by the fact that there are now n intensities instead of a finite number k; this is easily demonstrated for the M and U tests, although it seems difficult to construct a satisfactory argument for the g test. We may also note that, as was pointed out by Sargan (1953, p. 148), the intensity tests can be used when $\{X(t)\}$ is a discrete linear process. This follows at once, since with $X(t) = \sum_{u=-\infty}^{\infty} g(t-u)\,\epsilon(u)$ we have the relation (19), and by the central limit theorem, the distributions of $n^{-\frac{1}{2}} \sum_{r=-n}^{n} \epsilon(r) \cos \omega r$ and $n^{-\frac{1}{2}} \sum_{r=-n}^{n} \epsilon(r) \sin \omega r$ tend to the Normal form with zero mean and variance unity, at least when the third absolute moment $E\,|\,\epsilon(t)\,|^3$ is finite.

We could alternatively consider, instead of individual intensities, integrals of the form

$$K_n(\omega_1, \omega_2) = \int_{\omega_1}^{\omega_2} I_{2n+1}(\omega)\, d\omega = 2 \sum_{s=-2n}^{2n} \left(1 - \frac{|\,s\,|}{2n+1}\right) C_s (\sin \omega_2 s - \sin \omega_1 s)/s, \tag{20}$$

where $C_s = \sum_{r=1}^{2n+1-|s|} X(r)\,X(r+|s|)/(2n+1-|s|)$. Then for continuous $f_+(\omega)$ we have

$$\lim_{n\to\infty} E\{K_n(\omega_1,\omega_2)\} = 2\pi\int_{\omega_1}^{\omega_2} f_+(\omega)\,d\omega, \tag{21}$$

$$\lim_{n\to\infty} n \operatorname{var} K_n(\omega_1,\omega_2) = \pi\int_{\omega_1}^{\omega_2} \{2\pi f_+(\omega)\}^2\,d\omega \tag{22}$$

and

$$\lim_{n\to\infty} n \operatorname{cov}\{K_n(\omega_1,\omega_2), K_n(\omega_3,\omega_4)\} = 0, \tag{23}$$

when (ω_1,ω_2) and (ω_3,ω_4) do not overlap (see Grenander, 1951, pp. 521–3; Grenander & Rosenblatt, 1953, pp. 541–3). It can also be shown that under wide conditions (any which ensure the asymptotic normality of a finite set of serial covariances C_s will suffice) the distribution of a finite set of quantities such as $n^{\frac{1}{2}}[K_n(\omega_1,\omega_2) - E\{K_n(\omega_1,\omega_2)\}]$ tends, as $n\to\infty$, to the Normal form with zero means and finite covariance matrix. Hence, provided that

$$E\{K_n(\omega_i,\omega_{i+1})\} - 2\pi\int_{\omega_i}^{\omega_{i+1}} f_+(\omega)\,d\omega = o(n^{-\frac{1}{2}}) \quad \text{as} \quad n\to\infty,$$

which is certainly true if $f_+(\omega)$ has a bounded derivative (Grenander & Rosenblatt, 1953, Theorem 2), the statistics

$$v_i = n^{\frac{1}{2}}\left[K_n(\omega_i,\omega_{i+1}) - 2\pi\int_{\omega_i}^{\omega_{i+1}} f_+(\omega)\,d\omega\right] \Big/ \sqrt{\left[\pi\int_{\omega_i}^{\omega_{i+1}} \{2\pi f_+(\omega)\}^2\,d\omega\right]}, \tag{24}$$

where the intervals (ω_i,ω_{i+1}) do not overlap, are for large n distributed approximately as independent Normal variables with zero mean and variance unity.

This enables us to construct goodness of fit tests based on the v_i; for example, we might use the criterion $\sum_{i=0}^{k-1} v_i^2$, rejecting the null hypothesis if this is significantly large when referred to a χ^2 distribution with k degrees of freedom. Such tests should be much more powerful than those of § 3 because the asymptotic variance of K_n is not finite but proportional to n^{-1}. However, there is the disadvantage that when the null hypothesis does not specify $f_+(\omega)$ completely, the substitution of estimators for the unknown parameters may affect the asymptotic distribution of the test criteria. For the root-mean-square errors of these estimators will usually be of order $n^{-\frac{1}{2}}$, and an error of $n^{-\frac{1}{2}}$ in $f_+(\omega)$ produces an error in v_i which may be of order unity.

5. Numerical illustration

To illustrate the methods described in § 3, these were applied to test the goodness of fit of a set of observations from a particular series $\{X(t)\}$, a third-order linear autoregressive process with

$$X(t+2) - 1{\cdot}1X(t+1) + 0{\cdot}5X(t) = Y(t), \quad Y(t+1) - 0{\cdot}1\,Y(t) = \epsilon(t), \tag{25}$$

the $\epsilon(t)$ being independent Normal variables with zero mean and variance unity. The set consisted of 385 consecutive terms out of a group of 500 given by M. G. Kendall (1949, Series 16), n thus being 192. The end-points of the intervals on the ω axis were taken to be $\omega_i = \frac{1}{24}i\pi$ $(i = 0, 1, \ldots, 24)$ so that each interval was of length $\frac{1}{24}\pi = 8\pi/n$; this might be thought rather small, but an examination of the exact expressions for the variances and

covariances of the quantities $A_{2n}(\omega_{i+1})-A_{2n}(\omega_i)$, $B_{2n}(\omega_{i+1})-B_{2n}(\omega_i)$ along the lines of § 6 indicated that the χ^2 approximations to the distributions of the T_i should be fairly accurate.

From (25), we have

$$f_+(\omega) = 1/[\pi\,|\,(1-1{\cdot}1e^{i\omega}+0{\cdot}5e^{2i\omega})\,(1-0{\cdot}1e^{i\omega})\,|^2]$$
$$= 1/[\pi(2\cos^2\omega-3{\cdot}3\cos\omega+1{\cdot}46)\,(1{\cdot}01-0{\cdot}20\cos\omega)]. \qquad (26)$$

The graph of $f_+(\omega)$ plotted against ω is shown in Fig. 1; it has a single fairly sharp peak at a value of ω approximately equal to $\frac{1}{6}\pi$. Also on integrating (26) we find

$$\pi F_+(\omega) = 0{\cdot}05643\tan^{-1}\left(\frac{11x}{9}\right)+0{\cdot}79173\log\frac{2{\cdot}6(x+0{\cdot}34185)^2+0{\cdot}09617}{2{\cdot}6(x-0{\cdot}34185)^2+0{\cdot}09617}$$
$$+3{\cdot}33039\tan^{-1}\frac{0{\cdot}38465x}{0{\cdot}15385-x^2}, \qquad (27)$$

where $x=\tan(\frac{1}{2}\omega)$.

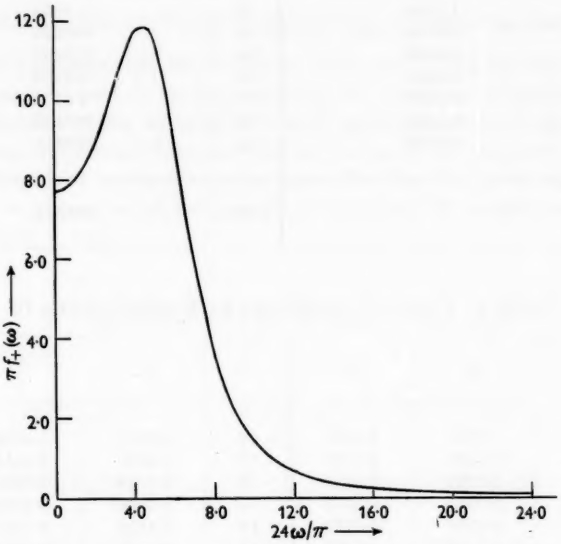


Fig. 1. Spectral density function for Kendall's Series 16.

The values of $\pi[F_+(\omega_{i+1})-F_+(\omega_i)]$ for $i=0,1,\ldots,23$ are given in Table 1. The corresponding values of

$$P_i = 4\{A_{2n}(\omega_{i+1})-A_{2n}(\omega_i)\}^2/\{F_+(\omega_{i+1})-F_+(\omega_i)\},$$
$$Q_i = 4\{B_{2n}(\omega_{i+1})-B_{2n}(\omega_i)\}^2/\{F_+(\omega_{i+1})-F_+(\omega_i)\},$$

and $T_i = P_i+Q_i$ are given in Table 2.

The likelihood ratio criterion $M=\sum_{i=0}^{23}\{T_i-2\log T_i\}+48(\log 2-1)=24{\cdot}34$, which is clearly not significantly large when referred to a χ^2 distribution with 24 degrees of freedom, so that with the M test we should draw the correct conclusion that the observations are consistent with the spectral distribution function of the series being the specified $F_+(\omega)$.

If the variance of $\epsilon(t)$ had not been specified, we would have used the homogeneity of variances criterion $M' = 23{\cdot}67$ with 23 degrees of freedom, which again is not significantly large (the difference $M - M'$, with 1 degree of freedom, which gives a test for $\sigma_\epsilon^2 = 1$ against alternatives $\sigma_\epsilon^2 \neq 1$, is also not significant).

Table 1. *Distribution of spectral 'power' over intervals of length* $\frac{1}{24}\pi$ *for Kendall's Series* 16

i	$\pi[F_+(\frac{1}{24}(i+1)\,\pi) - F_+(\frac{1}{24}i\pi)]$	i	$\pi[F_+(\frac{1}{24}(i+1)\,\pi) - F_+(\frac{1}{24}i\pi)]$
0	1·0220	12	0·0763
1	1·0953	13	0·0576
2	1·2440	14	0·0448
3	1·4421	15	0·0361
4	1·5558	16	0·0299
5	1·3799	17	0·0255
6	0·9734	18	0·0223
7	0·6029	19	0·0199
8	0·3668	20	0·0183
9	0·2309	21	0·0171
10	0·1523	22	0·0164
11	0·1056	23	0·0161
		Total	10·5513

Table 2. *Values of test statistics for Kendall's Series* 16

i	P_i	Q_i	T_i	i	P_i	Q_i	T_i
0	0·1949	2·2857	2·4806	12	0·0001	1·4326	1·4327
1	0·4505	0·6195	1·0700	13	0·0053	2·8015	2·8068
2	1·0072	2·8562	3·8634	14	0·1729	0·0048	0·1777
3	0·0134	0·0336	0·0470	15	1·7855	0·8365	2·6220
4	0·1471	0·6799	0·8270	16	0·4132	0·7761	1·1893
5	0·8415	0·1389	0·9804	17	0·7310	0·0200	0·7510
6	0·5004	4·8900	5·3904	18	0·5207	0·7617	1·2824
7	0·0069	0·2930	0·2999	19	0·1436	4·1721	4·3157
8	0·2438	2·5375	2·7813	20	0·0001	2·3556	2·3557
9	1·0291	0·4558	1·4849	21	0·0027	0·0886	0·0913
10	1·4156	0·5018	1·9174	22	0·6150	0·8077	1·4227
11	0·1182	0·1355	0·2537	23	0·0400	0·4936	0·5336
				Total	10·3987	29·9782	40·3769

Again we have

$$U = -2\sum_{i=0}^{23} \log 2y_i = \sum_{T_i \geqslant 2\log 2} T_i - \sum_{T_i < 2\log 2} 2\log(1 - e^{-\frac{1}{2}T_i}) - 48\log 2 = 40{\cdot}63,$$

with 48 degrees of freedom, and since there are 12 values of i with $T_i \geqslant 2\log 2$,

$$U_1 = \sum_{T_i \geqslant 2\log 2} T_i - 24\log 2 = 24{\cdot}39, \quad \text{with 24 degrees of freedom,}$$

and $$U_2 = \sum_{T_i < 2\log 2} \log(1 - e^{-\frac{1}{2}T_i}) - 24\log 2 = 16{\cdot}24, \quad \text{with 24 degrees of freedom.}$$

Neither U, U_1 nor U_2 is significantly large.

As a further check, values of the corresponding criteria U, U_1, U_2 for the sets $\{P_i\}$ and $\{Q_i\}$ of approximately independent variables each distributed as $\chi^2_{(1)}$, were obtained. For the P set, $U_1 = 41{\cdot}17\,(30\text{ D.F.})$, $U_2 = 7{\cdot}05\,(18\text{ D.F.})$, $U = 48{\cdot}22\,(48\text{ D.F.})$, and for the Q set, $U_1 = 13{\cdot}95\,(14\text{ D.F.})$, $U_2 = 30{\cdot}79\,(34\text{ D.F.})$, $U = 44{\cdot}74\,(48\text{ D.F.})$. The values for the Q set agree closely with their expectations, but for the P set, U_1 is greater than the upper 10% point of $\chi^2_{(30)}$, while U_2 is only just greater than the lower 1% point of $\chi^2_{(18)}$. This indicates that the values of P_i tend to be too low, which is confirmed by the fact that $\sum_{i=0}^{23} P_i = 10{\cdot}40$ is less than the lower 1% point of $\chi^2_{(24)}$. This effect is somewhat surprising. The exact expressions show that the values of $E(P_i)$ tend to be less than unity, but the negative bias does not seem to be large enough to provide a satisfactory explanation. A detailed examination of the accuracy of the χ^2 approximations in this particular case might be of interest, but has not been carried out because of the extremely tedious calculations required. There is of course always the possibility that the effect might be traced to some peculiarity in the series of residuals $\{\epsilon(t)\}$ (compare the experience of Bartlett & Rajalakshman, 1953, p. 120).

6. Derivation of the approximations of §2

It is easy to show that

$$4\pi^2 E[A_{2n}(\omega_2) - A_{2n}(\omega_1)]^2 = \int_0^\pi [K_n^{12}(u) - L_n^{12}(u)]^2\, dF_+(u), \tag{28}$$

$$4\pi^2 E[B_{2n}(\omega_2) - B_{2n}(\omega_1)]^2 = \int_0^\pi [K_n^{12}(u) + L_n^{12}(u)]^2\, dF_+(u), \tag{29}$$

$$4\pi^2 E[\{A_{2n}(\omega_4) - A_{2n}(\omega_3)\}\{A_{2n}(\omega_2) - A_{2n}(\omega_1)\}] = \int_0^\pi [K_n^{12}(u) - L_u^{12}(u)][K_n^{34}(u) - L_n^{34}(u)]\, dF_+(u)$$

and (30)

$$4\pi^2 E[\{B_{2n}(\omega_4) - B_{2n}(\omega_3)\}\{B_{2n}(\omega_2) - B_{2n}(\omega_1)\}] = \int_0^\pi [K_n^{12}(u) + L_n^{12}(u)][K_n^{34}(u) + L_n^{34}(u)]\, dF_+(u), \tag{31}$$

where $$K_n^{12}(u) = K_n(u, \omega_1, \omega_2) = G_n(u - \omega_1) - G_n(u - \omega_2), \tag{32}$$

$$L_n^{12}(u) = L_n(u, \omega_1, \omega_2) = G_n(u + \omega_1) - G_n(u + \omega_2), \tag{33}$$

G_n being defined by

$$\left.\begin{aligned} G_n(u) &= \int_0^u \frac{\sin(n + \frac{1}{2})y}{2\sin\frac{1}{2}y}\, dy, \\ \text{and} \quad K_n^{34}(u) &= K_n(u, \omega_3, \omega_4), \quad L_n^{34}(u) = L_n(u, \omega_3, \omega_4). \end{aligned}\right\} \tag{34}$$

For example,

$$4\pi^2 E[A_{2n}(\omega_2)-A_{2n}(\omega_1)]^2 = \sum_{r,s=-n}^{n}\left(\frac{\sin\omega_2 r-\sin\omega_1 r}{r}\right)\left(\frac{\sin\omega_2 s-\sin\omega_1 s}{s}\right)\int_0^\pi \cos u(r-s)\,dF_+(u)$$

$$= \int_0^\pi \left|\sum_{r=-n}^{n}[e^{iur}(e^{i\omega_2 r}-e^{-i\omega_2 r}-e^{i\omega_1 r}+e^{-i\omega_1 r})/(2r)]\right|^2 dF_+(u)$$

$$= \int_0^\pi [G_n(u+\omega_2)-G_n(u-\omega_2)-G_n(u+\omega_1)+G_n(u-\omega_1)]^2\,dF_+(u),$$

since $$\sum_{r=-n}^{n}(e^{iur}-1)/r = iu+2i\int_0^u\left(\sum_{r=1}^{n}\cos yr\right)dy = i\int_0^u \frac{\sin(n+\frac{1}{2})y}{\sin\frac{1}{2}y}\,dy. \tag{35}$$

(29), (30) and (31) are obtained in the same way.

Now $$G_n(u) = \int_0^u \frac{\sin(n+\frac{1}{2})y}{y}\,dy + \int_0^u \sin(n+\tfrac{1}{2})y\left\{\frac{1}{2\sin\frac{1}{2}y}-\frac{1}{y}\right\}dy$$

$$= Si(n+\tfrac{1}{2})u + I_n(u), \quad \text{say,} \tag{36}$$

where by the Riemann–Lebesgue theorem (see, for example, Titchmarsh, 1944, p. 403), $I_n(u)\to 0$ like n^{-1} as $n\to\infty$. In fact, on integrating the second term of (36) by parts, we find that for $|u|\leqslant\pi$,

$$|I_n(u)| \leqslant \frac{1}{n+\frac{1}{2}}\left(1-\frac{2}{\pi}\right) < 0{\cdot}37/n. \tag{37}$$

When the argument of G_n exceeds π, as can occur in (33), it is convenient to use the relation $G_n(2\pi-u) = \pi - G_n(u)$, which follows at once from (36), and to write

$$G_n(u) = \pi - Si(n+\tfrac{1}{2})(2\pi-u) + I_n(u),$$

where $I_n(u)$ still satisfies (37). Hence if

$$k_n^{12}(u) = k_n(u,\omega_1,\omega_2) = Si(n+\tfrac{1}{2})(u-\omega_1) - Si(n+\tfrac{1}{2})(u-\omega_2)$$

$$l_n^{12}(u) = l_n(u,\omega_1,\omega_2) = Si(n+\tfrac{1}{2})(u+\omega_1) - Si(n+\tfrac{1}{2})(u+\omega_2) \quad \text{for} \quad 0\leqslant u\leqslant\pi-\omega_2,$$

a similar definition applying for $u>\pi-\omega_2$ if $Si(n+\frac{1}{2})(u+\omega_i)$ is replaced by

$$\pi - Si(n+\tfrac{1}{2})(u+\omega_i) \quad \text{when} \quad u>\pi-\omega_i \quad (i=1,2),$$

$k_n^{12}(u)$ and $l_n^{12}(u)$ will be, for moderately large n, good approximations to $K_n^{12}(u)$ and $L_n^{12}(u)$ respectively, and similarly $k_n^{34}(u)$ and $l_n^{34}(u)$ will be good approximations to $K_n^{34}(u)$ and $L_n^{34}(u)$.

Now $Si(n+\frac{1}{2})u$ may be replaced to a good approximation by $\frac{1}{2}\pi$ when $u>\lambda\pi/(n+\frac{1}{2})$, $\lambda = O(1)$. In fact, since

$$\int_y^\infty \frac{\sin u}{u}\,du = \frac{\cos y}{y}+\frac{\sin y}{y^2}-\int_y^\infty \frac{2\sin u}{u^3}\,du = \frac{\cos y}{y}+\frac{2\theta_y}{y^2} \quad \text{with} \quad |\theta_y|<1,$$

we have $$|Si(n+\tfrac{1}{2})u - \tfrac{1}{2}\pi| < \frac{1}{\lambda\pi}+\frac{2}{\lambda^2\pi^2}, \quad u>\lambda\pi/(n+\tfrac{1}{2}),$$

so that with $\lambda=2$, for example, the error in the approximation is not more than about 10 %. It follows that to a good approximation we can put

$$l_n^{12}(u) = 0, \quad (0\leqslant u\leqslant\pi), \quad \text{provided that} \quad \omega_1 \gg \pi/n, \quad \pi-\omega_2 \gg \pi/n, \tag{38}$$

and $$k_n^{12}(u)\begin{cases} = 0, & u<\omega_1-\lambda\pi/n, \quad u>\omega_2+\lambda\pi/n, \\ = \pi, & \omega_1+\lambda\pi/n\leqslant u\leqslant\omega_2-\lambda\pi/n \end{cases} \quad (\text{for } \omega_2-\omega_1>2\lambda\pi/n). \tag{39}$$

Hence (28) and (29) are both approximately equal to

$$\int_0^\pi \{k_n^{12}(u)\}^2\, dF_+(u),$$

and therefore to

$$\pi^2[F_+(\omega_2-\lambda\pi/n)-F_+(\omega_1+\lambda\pi/n)]+\sum_{i=1}^{2}\int_{\omega_i-\lambda\pi/n}^{\omega_i+\lambda\pi/n}\{k_n^{12}(u)\}^2\, dF_+(u),$$

from which the result (11) of § 2 follows under the stated conditions.

Also when approximations similar to (38) and (39) can be used for $l_n^{34}(u)$ and $k_n^{34}(u)$, (30) and (31) are approximately equal to $\int_0^\pi k_n^{12}(u)\, k_n^{34}(u)\, dF_+(u)$, and, using Schwarz's inequality,

$$\left|\int_{\omega_3+\lambda\pi/n}^{\pi} k_n^{12}(u)\, k_n^{34}(u)\, dF_+(u)\right|,\quad \left|\int_0^{\omega_1-\lambda\pi/n} k_n^{12}(u)\, k_n^{34}(u)\, dF_+(u)\right| \ll \sqrt{(V_{12}V_{34})},$$

where

$$V_{12}=\int_0^\pi \{k_n^{12}(u)\}^2\, dF_+(u),\quad V_{34}=\int_0^\pi \{k_n^{34}(u)\}^2\, dF_+(u),$$

so that for $\omega_3-\omega_2>2\lambda\pi/n$,

$$\left|\int_0^\pi k_n^{12}(u)\, k_n^{34}(u)\, dF_+(u)\right| \ll \sqrt{(V_{12}V_{34})}. \tag{40}$$

(40) will also hold for $\omega_3-\omega_2<2\lambda\pi/n$ if

$$\left|\int_{\omega_3-\lambda\pi/n}^{\omega_3+\lambda\pi/n} k_n^{12}(u)\, k_n^{34}(u)\, dF_+(u)\right| \ll \sqrt{(V_{12}V_{34})},$$

which is certainly true when

$$\int_{\omega_3-\lambda\pi/n}^{\omega_3+\lambda\pi/n}\{k_n^{12}(u)\}^2\, dF_+(u)\ll V_{12},\quad \int_{\omega_3-\lambda\pi/n}^{\omega_3+\lambda\pi/n}\{k_n^{34}(u)\}^2\, dF_+(u)\ll V_{34}.$$

The results (12) of § 2 then follow.

The use of (11) and (12) when the end-point of an interval coincides with 0 or π can be justified by a similar argument.

To make the above discussion rigorous we require expressions for upper bounds to the errors in the approximations. These are easily obtained, although their form is somewhat complicated. For example

$$4\pi^2 E\{A_{2n}(\omega_2)-A_{2n}(\omega_1)\}^2=\int_0^\pi (k_n(u)-l_n(u)+\delta_n(u))^2\, dF_+(u) \tag{41}$$

(dropping the superscripts from k and l without ambiguity), where from (37),

$$|\,\delta_n(u)\,|<1{\cdot}5/n,\quad (0\leqslant u\leqslant\pi).$$

Now (41) is equal to

$$\int_0^\pi k_n^2(u)\, dF_+(u)+R_n^{(1)}+R_n^{(2)}, \tag{42}$$

where

$$R_n^{(1)}=\int_0^\pi \delta_n^2(u)\, dF_+(u)+2\int_0^\pi \delta_n(u)\,(k_n(u)-l_n(u))\, dF_+(u) \tag{43}$$

and

$$R_n^{(2)}=\int_0^\pi l_n^2(u)\, dF_+(u)-2\int_0^\pi l_n(u)\, k_n(u)\, dF_+(u). \tag{44}$$

Also
$$\int_0^\pi k_n^2(u)\,dF_+(u) = \pi^2\{F_+(\omega_2) - F_+(\omega_1)\} + R_n^{(3)} + R_n^{(4)}, \tag{45}$$

where
$$R_n^{(3)} = \left(\int_0^{\omega_1 - \lambda\pi/(n+\frac{1}{2})} + \int_{\omega_2+\lambda\pi/(n+\frac{1}{2})}^{\pi}\right) k_n^2(u)\,dF_+(u) + \int_{\omega_1+\lambda\pi/(n+\frac{1}{2})}^{\omega_2-\lambda\pi/(n+\frac{1}{2})} (k_n^2(u) - \pi^2)\,dF_+(u) \tag{46}$$

and
$$R_n^{(4)} = \left(\int_{\omega_1-\lambda\pi/(n+\frac{1}{2})}^{\omega_1} + \int_{\omega_2}^{\omega_2+\lambda\pi/(n+\frac{1}{2})}\right) k_n^2(u)\,dF_+(u) + \left(\int_{\omega_1}^{\omega_1+\lambda\pi/(n+\frac{1}{2})} + \int_{\omega_2-\lambda\pi/(n+\frac{1}{2})}^{\omega_2}\right)(k_n^2(u) - \pi^2)\,dF_+(u)$$
$$= \int_{-\lambda\pi/(n+\frac{1}{2})}^{\lambda\pi/(n+\frac{1}{2})} W_n(\omega_2 - \omega_1, u)\,\{dF_+(\omega_1 + u) - dF_+(\omega_2 - u)\}, \tag{47}$$

where
$$W_n(\omega_2 - \omega_1, u) \begin{cases} = k_n^2(\omega_1 + u) = k_n^2(\omega_2 - u) & (u \leqslant 0), \\ = k_n^2(\omega_1 + u) - 1 & (u > 0). \end{cases} \tag{48}$$

Hence the magnitude of the relative error in the approximation does not exceed
$$\sum_{i=1}^{4} |R_n^{(i)}| / [\pi^2\{F_+(\omega_2) - F_+(\omega_1)\}]. \tag{49}$$

Upper bounds for the first three remainder terms (43), (44) and (46) may be derived by using the inequality
$$|Si(n+\tfrac{1}{2})u - \tfrac{1}{2}\pi| < \frac{1}{u} + \frac{2}{u^2} = \phi(u), \text{ say} \quad (u > 0).$$

This gives
$$|l_n(u)| < \sum_{i=1}^{2} \phi[(n+\tfrac{1}{2})\min(\omega_i + u, 2\pi - \omega_i - u)] < L \quad (0 \leqslant u \leqslant \pi), \tag{50}$$

where
$$L = \frac{2}{(n+\frac{1}{2})\min(\omega_1, \pi - \omega_2)}\left[1 + \frac{2}{(n+\frac{1}{2})\min(\omega_1, \pi - \omega_2)}\right], \tag{51}$$

$$|k_n(u)| < \sum_{i=1}^{2} \phi[(n+\tfrac{1}{2})(\omega_i - u)] < K_1, \quad u < \omega_1 - \lambda\pi/(n+\tfrac{1}{2}), \tag{52}$$

where
$$K_1 = 4/(3\lambda\pi) + 20/(9\lambda^2\pi^2) \tag{53}$$

(assuming that $\omega_2 - \omega_1 > 2\lambda\pi/n$),

$$|k_n(u)| < \sum_{i=1}^{2} \phi[(n+\tfrac{1}{2})(u - \omega_1)] < K_1, \quad u > \omega_2 + \lambda\pi/(n+\tfrac{1}{2}) \tag{54}$$

and
$$|k_n(u) - \pi| < \phi[(n+\tfrac{1}{2})(u - \omega_1)] + \phi[(n+\tfrac{1}{2})(\omega_2 - u)] < K_2, \quad \omega_1 + \lambda\pi/(n+\tfrac{1}{2}) < u < \omega_2 - \lambda\pi/(n+\tfrac{1}{2}), \tag{55}$$

where
$$K_2 = 2/\lambda\pi + 4/\lambda^2\pi^2. \tag{56}$$

Hence we have
$$|R_n^{(1)}| < (1{\cdot}5/n)^2 F_+(\pi) + (3/n)\left[LF_+(\pi) + \int_0^\pi |k_n(u)|\,dF_+(u)\right], \tag{57}$$

$$|R_n^{(2)}| < L^2 F_+(\pi) + 2L\int_0^\pi |k_n(u)|\,dF_+(u), \tag{58}$$

with
$$\int_0^\pi |k_n(u)|\,dF_+(u) < 1{\cdot}2\pi[F_+(\omega_2 + \lambda\pi/(n+\tfrac{1}{2})) - F_+(\omega_1 - \lambda\pi/(n+\tfrac{1}{2}))] + K_1 F_+(\pi)$$
$$(|k_n(u)| < 1{\cdot}2\pi \text{ for all } u), \tag{59}$$

and
$$|R_n^{(3)}| < K_1^2[F_+(\pi) - F_+(\omega_2 + \lambda\pi/(n+\tfrac{1}{2})) + F_+(\omega_1 - \lambda\pi/(n+\tfrac{1}{2}))]$$
$$+ K_2(K_2+2\pi)[F_+(\omega_2 - \lambda\pi/(n+\tfrac{1}{2})) - F_+(\omega_1 + \lambda\pi/(n+\tfrac{1}{2}))]$$
$$< K_1^2 F_+(\pi) + K_2(K_2+2\pi)[F_+(\omega_2) - F_+(\omega_1)]. \quad (60)$$

For the fourth remainder term (47), we note that to a good approximation W_n is independent of $\omega_2 - \omega_1$, increasing from 0 to $\frac{1}{4}\pi^2$ as u increases from $-\lambda\pi/(n+\frac{1}{2})$ to 0, and from $-\frac{3}{4}\pi^2$ to 0 as u increases from 0 to $\lambda\pi/(n+\frac{1}{2})$. Hence $R_n^{(4)}$ is at most of the order of

$$\left|\sum_{i=1}^{2}[F_+(\omega_i + \lambda\pi/n) - F_+(\omega_i - \lambda\pi/n)]\right|.$$

In fact we have
$$0 < W_n(\omega_2 - \omega_1, u) < \{\tfrac{1}{2}\pi + \phi(\lambda\pi)\}^2 \quad (u \leqslant 0),$$
$$0 > W_n(\omega_2 - \omega_1, u) > \{\tfrac{1}{2}\pi - \phi(\lambda\pi)\}^2 - \pi^2 \quad (u > 0)$$

(provided that $\frac{1}{2}\pi > \phi(\lambda\pi)$, which certainly holds for $\lambda > 1$), so that

$$|R_n^{(4)}| < \max[\{\tfrac{1}{2}\pi + \phi(\lambda\pi)\}^2\{F_+(\omega_1) - F_+(\omega_1 - \lambda\pi/(n+\tfrac{1}{2})) + F_+(\omega_2 + \lambda\pi/(n+\tfrac{1}{2})) - F_+(\omega_2)\},$$
$$\{\pi^2 - (\tfrac{1}{2}\pi - \phi(\lambda\pi))^2\}\{F_+(\omega_1 + \lambda\pi/(n+\tfrac{1}{2})) - F_+(\omega_1) + F_+(\omega_2) - F_+(\omega_2 - \lambda\pi/(n+\tfrac{1}{2}))\}], \quad (61)$$

or, when $\phi(\lambda\pi)$ can be neglected in comparison with $\frac{1}{2}\pi$,

$$|R_n^{(4)}| < \pi^2 \max[\tfrac{1}{4}\{F_+(\omega_1) - F_+(\omega_1 - \lambda\pi/n) + F_+(\omega_2 + \lambda\pi/n) - F_+(\omega_2)\},$$
$$\tfrac{3}{4}\{F_+(\omega_1 + \lambda\pi/n) - F_+(\omega_1) + F_+(\omega_2) - F_+(\omega_2 - \lambda\pi/n)\}]. \quad (62)$$

From (57)–(62) we obtain an upper bound, U say, for (49). For U to be small, we certainly must take λ to be appreciably greater than unity, since the term $K_2(K_2+2\pi)/\pi^2 > 2/(5\lambda)$ occurs; the value of this for $\lambda = 4$, for example, is 0·12. However, λ must also not be too large because of the contribution to U from (61) or (62), which in general will be small only for $\omega_2 - \omega_1 \gg \lambda\pi/n$. Thus the condition that U is to be small may impose a fairly severe restriction on the minimum admissible length of interval $\omega_2 - \omega_1$. The contribution from the remaining terms in U, which is equal to

$$\{C_1 F_+(\pi) + C_2[F_+(\omega_2 + \lambda\pi/(n+\tfrac{1}{2})) - F_+(\omega_1 - \lambda\pi/(n+\tfrac{1}{2}))]\}/\{F_+(\omega_2) - F_+(\omega_1)\}, \quad (63)$$

where
$$C_1 = [(L + 1{\cdot}5/n)^2 + K_1(K_1 + 2L + 3/n)]/\pi^2, \quad C_2 = 1{\cdot}2\,(2L + 3/n)/\pi,$$

will usually be negligible unless $F_+(\omega_2) - F_+(\omega_1)$ is a very small fraction of $F_+(\pi)$; e.g. if $\lambda \geqslant 4$, ω_1 and $\pi - \omega_2 \geqslant 16\pi/n$, $n = 200$, we have $C_1 < 0{\cdot}003$ and $C_2 < 0{\cdot}04$.

From the behaviour of the functions $W_n(u)$, $l_n(u)$ and $k_n(u)$ it is easy to see that U will often be a very conservative upper bound, so that the approximation may be quite accurate even when U is not small (this is the case, for instance, in the example of § 5). However, to obtain a better estimate of the relative error it seems to be necessary to make some detailed assumptions about the behaviour of $F_+(\omega)$. For example, if $F_+(\omega)$ has a derivative $f_+(\omega)$ which is continuous over an interval covering (ω_1, ω_2), we can show that the approximation may be quite good even when $\omega_2 - \omega_1$ is only of order π/n. In fact if we have

$$\left|f_+(\omega) - f_+\left(\frac{\omega_1+\omega_2}{2}\right)\right| < \varepsilon \quad \text{for} \quad \left|\omega - \frac{\omega_1+\omega_2}{2}\right| < \delta,$$

where $\delta \geqslant (\lambda+\frac{1}{2}\mu)\pi/(n+\frac{1}{2})$, with $\mu=(n+\frac{1}{2})(\omega_2-\omega_1)/\pi$, and assume that ϵ may be neglected, we find that the contribution to the relative error of $R_n^{(4)}$ and the last integral in $R_n^{(3)}$, from which the dominant terms in U usually arise, is

$$(2/\mu\pi^3)\int_0^{\lambda+\frac{1}{2}\mu}\{Si(u+\tfrac{1}{2}\mu\pi)-Si(u-\tfrac{1}{2}\mu\pi)\}^2\,du, \tag{64}$$

which is, for example, $<0{\cdot}10$ when $\mu>4$, $\lambda>2$. [This follows quite simply by using the results

$$\int_0^y S(u)\,du=1+2\theta_1/y, \quad \int_0^y S^2(u)\,du=\tfrac{1}{2}\pi-1/2y+\theta_2/y^2 \quad (|\theta_1|, |\theta_2|<1),$$

where $S(u)=\frac{1}{2}\pi-Si(u)$.]

All the above applies equally to the approximation to $4\pi^2E\{B_{2n}(\omega_2)-B_{2n}(\omega_1)\}^2$, since we only have to change the sign of $l_n(u)$ in (41). The expressions for the covariances, given by (30) and (31), may be dealt with in much the same way; we may note that for $\omega_2=\omega_3$, it can be shown, by considering an integral similar to (64), that taking these to be zero may be a good approximation for $\omega_2-\omega_1$, $\omega_4-\omega_2$ of order π/n when $f_+(\omega)$ can be treated as constant over an interval containing ω_2.

APPENDIX

Asymptotic power functions for the tests of § 3

(1) *The U test*

We assume that to a sufficiently good approximation, $T_i=2\sigma_i^2x_i$, where the x_i are distributed independently with probability density functions e^{-x_i} $(0<x_i<\infty)$. The y_i will then be distributed independently with probability density functions

$$(1+\epsilon_i)(y_i^{\epsilon_i}+(1-y_i)^{\epsilon_i}) \quad (0<y_i<\tfrac{1}{2};\quad i=0,1,\ldots,k-1), \tag{1}$$

where $\epsilon_i=(1/\sigma_i^2)-1$.

Let $v_i=\frac{1}{2}u_i=-\log 2y_i$. Then from (1) we have

$$\begin{aligned}E(v_i^r)&=(1+\epsilon_i)\left[(\tfrac{1}{2})^{1+\epsilon_i}\int_0^\infty v_i^r e^{-v_i(1+\epsilon_i)}\,dv_i+\frac{1}{2}\int_0^\infty v_i^r e^{-v_i}(1-\tfrac{1}{2}e^{-v_i})^{\epsilon_i}\,dv_i\right]\\
&=(1+\epsilon_i)^{-r}(\tfrac{1}{2})^{1+\epsilon_i}\,\Gamma(r+1)\\
&\qquad+\tfrac{1}{2}(1+\epsilon_i)\int_0^\infty v_i^r e^{-v_i}\left(1+\sum_{s=1}^{\infty}(-)^s\epsilon_i(\epsilon_i-1)\ldots(\epsilon_i-s+1)(\tfrac{1}{2}e^{-v_i})^s/s!\right)dv_i\\
&=\tfrac{1}{2}(1+\epsilon_i)\,\Gamma(r+1)\left[(\tfrac{1}{2})^{\epsilon_i}(1+\epsilon_i)^{-(r+1)}+1+\sum_{s=1}^{\infty}(-)^s\frac{\epsilon_i(\epsilon_i-1)\ldots(\epsilon_i-s+1)}{s!\,2^s(s+1)^{r+1}}\right]\end{aligned} \tag{2}$$

(the interchange of the order of integration and summation being easily justified). Let the σ_i^2 have a positive lower bound and a finite upper bound. Then from (2) we easily see that the third absolute moments of the v_i about their means have a finite upper bound and that the variances of the v_i have a positive lower bound. Hence by Lyapunov's form of the central limit theorem (Cramér, 1946, p. 215), the distribution of $\{V-E(V)\}/\sqrt{(\text{var } V)}$, where $V=\sum_{i=0}^{k-1}v_i=\frac{1}{2}U$, tends to the Normal form with zero mean and variance unity when $k\to\infty$.

Let V_α be the critical value of V when the significance level of the test is α. Then for large k, $V_\alpha \doteqdot k+\xi_\alpha\sqrt{k}$, since $E(V)=k=\text{var } V$ when $\epsilon_i=0$, and the power function $\Pr(V>V_\alpha)$ is approximately equal to

$$1-\Phi[(k+\xi_\alpha\sqrt{k}-E(V))/\sqrt{(\text{var } V)}]. \tag{3}$$

Now when the ϵ_i are of order $k^{-\frac{1}{2}}$, (2) gives

$$\begin{aligned}E(V)&=k+\frac{1}{2}\sum_{i=0}^{k-1}\epsilon_i\left\{\log\tfrac{1}{2}-\sum_{s=1}^{\infty}\frac{1}{s(s+1)^2\,2^s}\right\}+O(1)\\&=k-0{\cdot}42\sum_{i=0}^{k-1}\epsilon_i+O(1),\end{aligned}$$

and var $V=k+O(\sqrt{k})$. Hence (3) then becomes approximately equal to

$$1-\Phi\left(\xi_\alpha+0{\cdot}42\,k^{-\frac{1}{2}}\sum_{i=0}^{k-1}\epsilon_i\right). \tag{4}$$

(2) *The modified U test*

The probability density function of y_i conditional on $p_i<\frac{1}{2}$ is

$$(\tfrac{1}{2})^{-(1+\epsilon_i)}(1+\epsilon_i)\,y_i^{\epsilon_i}\quad(0<y_i<\tfrac{1}{2}).$$

Hence

$$E(v_i^r\,|\,p_i<\tfrac{1}{2})=(1+\epsilon_i)^{-r}\,\Gamma(r+1). \tag{5}$$

Similarly we have

$$E(v_i^r\,|\,p_i\geqslant\tfrac{1}{2})=\tfrac{1}{2}(1+\epsilon_i)\,\Gamma(r+1)\left(1+\sum_{s=1}^{\infty}(-)^s\frac{\epsilon_i(\epsilon_i-1)\ldots(\epsilon_i-s+1)}{s!\,2^s(s+1)^{r+1}}\right)\Big/(1-(\tfrac{1}{2})^{1+\epsilon_i}). \tag{6}$$

It then follows, by the central limit theorem, that when the σ_i^2 have a positive lower bound and a finite upper bound, the distributions of $\{U_j-E(U_j)\}/\sqrt{(\text{var } U_j)}$ $(j=1,\ 2)$, conditional on k_1 specified values of i with $p_i\geqslant\frac{1}{2}$, tend to the Normal form with zero mean and variance unity when k_1 and $k_2=k-k_1\to\infty$.

Let k be large and the ϵ_i be of order $k^{-\frac{1}{2}}$. Since

$$\Pr(p_i<\tfrac{1}{2})=(\tfrac{1}{2})^{1+\epsilon_i}=\tfrac{1}{2}(1+\epsilon_i\log\tfrac{1}{2})+O(k^{-1}),$$

we have

$$E(k_1)=\tfrac{1}{2}k+O(\sqrt{k}),\quad \text{var}\,k_1=\tfrac{1}{4}k+O(1),$$

so that k_1, k_2 are $O(k)$. The conditional probabilities $\pi_j=\Pr\{U_j>2(k_j+\lambda\sqrt{k_j})\}$ are then given to a good approximation by

$$\left.\begin{aligned}\pi_1&=1-\Phi(\lambda-0{\cdot}16k_1^{-\frac{1}{2}}\Sigma_1\epsilon_i),\\ \pi_2&=1-\Phi(\lambda+k_2^{-\frac{1}{2}}\Sigma_2\epsilon_i)\end{aligned}\right\} \tag{7}$$

(Σ_1, Σ_2 having the same meaning as in § 3), since from (5) and (6),

$$E(\Sigma_1 v_i)=k_1+0{\cdot}16\,\Sigma_1\epsilon_i+O(1),\quad \text{var}\,\Sigma_1 v_i=k_1+O(\sqrt{k}),$$

and

$$E(\Sigma_2 v_i)=k_2-\Sigma_2\epsilon_i+O(1),\quad \text{var}\,\Sigma_2 v_i=k_2+O(\sqrt{k}).$$

Now let $2(k_j+\lambda\sqrt{k_j})$ be the critical value of U_j for the modified U test. From (7) with $\epsilon_i=0$, $\lambda\doteqdot\Phi^{-1}\{(1-\alpha)^{-\frac{1}{2}}\}$, so that when the significance level $\alpha\ll 1$, $\lambda\doteqdot\xi_{\frac{1}{2}\alpha}$. The power function of the test is

$$E(\pi_1+\pi_2-\pi_1\pi_2), \tag{8}$$

where the expectation is taken over all possible sets of values of i. Although (8) is a very complicated expression, its asymptotic form is easily found. For

$$E(\Sigma_2 \epsilon_i) = \sum_{i=0}^{k-1} (\tfrac{1}{2})^{1+\epsilon_i} \epsilon_i = \frac{1}{2} \sum_{i=0}^{k-1} \epsilon_i + O(1),$$

and

$$\operatorname{var} \Sigma_2 \epsilon_i = \sum_{i=0}^{k-1} (\tfrac{1}{2})^{1+\epsilon_i} \{1 - (\tfrac{1}{2})^{1+\epsilon_i}\} \epsilon_i^2 = \frac{1}{4} \sum_{i=0}^{k-1} \epsilon_i^2 + O(k^{-1}),$$

so that

$$\pi_1 \doteqdot 1 - \Phi(\lambda - 0{\cdot}16\theta/\sqrt{2}), \quad \pi_2 \doteqdot 1 - \Phi(\lambda + \theta/\sqrt{2}),$$

where $\theta = k^{-\frac{1}{2}} \sum_{i=0}^{k-1} \epsilon_i$. Hence (8) is approximately equal to

$$1 - \Phi(\lambda - 0{\cdot}16\theta/\sqrt{2})\, \Phi(\lambda + \theta/\sqrt{2}). \tag{9}$$

Clearly (9) approaches unity when θ takes sufficiently large positive or negative values.

(3) *The M test*

With $s_i^2 = \sigma_i^2 x_i$, the cumulant generating function of M becomes

$$\begin{aligned} \log E(\exp M\phi) &= \sum_{i=0}^{k-1} \log \left[e^{-2\phi} \int_0^\infty e^{-x(1-2\phi\sigma_i^2)} (\sigma_i^2 x)^{-2\phi}\, dx \right] \\ &= k(\log \Gamma(1-2\phi) - 2\phi) - \sum_{i=0}^{k-1} [2\phi \log \sigma_i^2 + (1-2\phi) \log (1 - 2\phi\sigma_i^2)] \\ &= k \left[\sum_{r=1}^{\infty} \frac{(-2\phi)^r \psi^{(r-1)}(1)}{r!} - 2\phi \right] + \sum_{i=0}^{k-1} \left[(1-2\phi) \sum_{r=1}^{\infty} \frac{(2\phi\sigma_i^2)^r}{r} - 2\phi \log \sigma_i^2 \right], \end{aligned}$$

where $\psi^{(r)}(x) = \left(\frac{d}{dx}\right)^{r-1} \log \Gamma(x)$. Hence the cumulants of M are given by

$$\kappa_1(M) = -2k\psi(1) + 2 \sum_{i=0}^{k-1} (\sigma_i^2 - 1 - \log \sigma_i^2), \tag{10}$$

$$\kappa_r(M) = (-2)^r k \psi^{(r-1)}(1) + 2^r (r-2)! \sum_{i=0}^{k-1} (\sigma_i^2)^{r-1} \{(r-1)\sigma_i^2 - r\} \quad (r > 1). \tag{11}$$

It is easily seen that when the σ_i^2 and their reciprocals have a finite upper bound, the distribution of M tends to the Normal form as $k \to \infty$. Also if the ϵ_i are of order $k^{-\frac{1}{2}}$, we have from (10) and (11),

$$E(M) = 1{\cdot}15k + \sum_{i=0}^{k-1} \epsilon_i^2 + O(k^{\frac{1}{2}}),$$

and

$$\operatorname{var} M = 2{\cdot}58k + O(1).$$

Thus for large k, $M_\alpha \doteqdot 1{\cdot}15k + \xi_\alpha \sqrt{(2{\cdot}58k)}$, M_α being the critical value of M for significance level α, and the power function $\Pr(M > M_\alpha)$ is approximately equal to

$$1 - \Phi\left\{ \xi_\alpha - \sum_{i=0}^{k-1} \epsilon_i^2 / \sqrt{(2{\cdot}58k)} \right\}. \tag{12}$$

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
BARTLETT, M. S. (1950). *Biometrika*, **37**, 1.
BARTLETT, M. S. (1954). *Publ. Inst. Statist. Univ. Paris*, **3**, Fasc. 3, p. 119.
BARTLETT, M. S. (1955). *Stochastic Processes*. Cambridge University Press.
BARTLETT, M. S. & RAJALAKSHMAN, D. V. (1953). *J. R. Statist. Soc.* B, **15**, 107.
BOX, G. E. P. (1949). *Biometrika*, **36**, 317.
CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton.
DOOB, J. L. (1952). *Stochastic Processes*. New York: Wiley.
FISHER, R. A. (1929). *Proc. Roy. Soc.* A, **125**, 54.
FISHER, R. A. (1941). *Statistical Methods for Research Workers*, 8th ed. Edinburgh: Oliver and Boyd.
GRENANDER, U. (1951). *Ark. Mat.* **1**, 503.
GRENANDER, U. & ROSENBLATT, M. (1953). *Ann. Math. Statist.* **24**, 537.
KENDALL, M. G. (1949). *Biometrika*, **36**, 267.
SARGAN, J. D. (1953). *J. R. Statist. Soc.* B, **15**, 140.
TITCHMARSH, E. C. (1944). *Theory of Functions*, 2nd ed. Oxford University Press.
WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala.
WHITTLE, P. (1952*a*). *Trab. Estadistica*, **3**, 43.
WHITTLE, P. (1952*b*). *Biometrika*, **39**, 309.
WHITTLE P. (1954). Appendix to *A Study in the Analysis of Stationary Time Series*, by H. Wold, 2nd ed. Uppsala.

SUFFICIENCY CONDITIONS IN REGULAR MARKOV CHAINS AND CERTAIN RANDOM WALKS

By J. GANI

*Nuffield Fellow, Statistical Laboratory, The University of Manchester**

For positively regular Markov chains with a finite number of states, transition probabilities of the form

$$p_{ij}(\theta) = \alpha_{ij} \exp \{K_{ij}\Lambda_1(\theta) + \lambda_2(\theta)\},$$

are known to admit a sufficient estimator of θ in realizations of the chain starting with a fixed state and consisting of a fixed number of transitions.

This paper considers whether transition probabilities of the same form will admit a sufficient estimator of θ in other finite regular, but not positively regular, Markov chains. For chains with an irreducible subset of two or more states, in which a realization starts from a fixed state and consists of a fixed number of transitions, these probabilities are found to admit a maximum-likelihood estimator of the function $g(\theta) = -\lambda_2'(\theta)/\Lambda_1'(\theta)$, which is sufficient and unbiased.

There is some difference in chains with an absorbing state, in which realizations start from a fixed state but continue until the absorbing state is reached; in this sequential case, the maximum-likelihood estimator, with the number of transitions in the realization, together provide a sufficient estimator of the function $g(\theta)$, which in general is no longer unbiased.

We restrict ourselves to the particular case where a certain linear relation is satisfied. This gives rise to some simple stochastic matrices admitting a sufficient estimator of θ, which consist of probabilities of the forms θ and $1-\theta$; in some of these cases, unbiased sufficient estimators of θ reduce to known results of Girshick, Mosteller & Savage (1946).

Some non-regular finite and infinite chains with absorbing states, associated with random walks, whose matrices consist of various patterns of probabilities θ and $1-\theta$, are also found to admit a sufficient estimator of θ. The paper ends with the examination of such an example, arising in sequential estimation.

1. Introduction

In a recent paper (Gani, 1955), the author considered some sufficiency conditions holding for positively regular Markov chains with a finite number s of states $E_1, \ldots, E_s$; for these chains, stationary probabilities exist and are all positive.

It was shown that, for a realization of the chain starting with a fixed initial state and consisting of a fixed number n of transitions, a sufficient estimator of θ, the parameter defining the transition probabilities $p_{ij}(\theta) = \Pr\{E_j \mid E_i\}$ $(i, j = 1, \ldots, s)$, existed if in any row i of the stochastic matrix $\mathbf{p} = \{p_{ij}\}$, these probabilities were of the form

$$p_{ij}(\theta) = \alpha_{ij} \exp \{K_{ij}\Lambda_1(\theta) + \lambda_2(\theta)\} \quad (j = 1, \ldots, s), \tag{1}$$

where the constants $\alpha_{ij} \geqslant 0$, and the number r of distinct exponents K_{ij} in the row could be less than or equal to the number s of states. Since $\sum_{j=1}^{s} p_{ij} = 1$, the functions $\Lambda_1(\theta)$ and $\lambda_2(\theta)$ were related by the equation

$$\exp \{-\lambda_2(\theta)\} = \sum_{j=1}^{s} \alpha_{ij} \exp \{K_{ij}\Lambda_1(\theta)\}; \tag{2}$$

from this it followed that the transition probabilities in the remaining rows of the matrix, except possibly for coefficients, were also given by the r distinct forms of the $p_{ij}(\theta)$ corresponding to the r distinct values of the K_{ij} in row i.

* The greater part of this work was completed while the author was at the Australian National University, Canberra A.C.T., Australia.

The question arises whether a form similar to (1), for the transition probabilities admitting a sufficient estimator of θ, exists when a chain is regular, that is, with stationary probabilities some of which may be zero. It is known (Bartlett, 1955, § 2·21; Feller, 1950, § 15·5) that regular chains, other than those which are positively regular, have stochastic matrices the simplest of which may be written as

$$\mathbf{p} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{Q} & \mathbf{R} \end{pmatrix}, \tag{3}$$

where $\mathbf{S}$, a square submatrix of transition probabilities depending on θ, constitutes a single irreducible closed subset, such that the states in it form a positively regular chain. The submatrices $\mathbf{Q}$ and $\mathbf{R}$ also consist of transition probabilities depending on θ, and $\mathbf{0}$ is a submatrix of zero elements. The case of particular interest, which arises in sequential problems, is that in which the irreducible closed set consists of a single 'absorbing' state permitting entry but no exit; $\mathbf{S}$ is then the single element 1.

In the two cases where the closed set consists (i) of two or more states none of which is closed, or (ii) of a single absorbing state, we shall consider two distinct types of realizations of the chains. For the first, exactly as with positively regular chains, we shall take realizations consisting of a fixed number n of transitions and starting from a fixed initial state; these may end with any one of the accessible states in the chain, nor does the possible entry of the system into the closed set terminate the process of transition from one state to another of it. For the second case, however, a realization starting from a fixed initial state and consisting of a similar fixed number n of transitions could result in the absorbing state being reached in $x \leqslant n$ transitions. In such cases, as is usual in sequential problems, we consider realizations in which the process is allowed to run until the absorbing state is reached; the number of transitions x required for this to happen is then itself a random variable. We proceed to examine these cases in greater detail.

2. Chains with a single irreducible closed subset of two or more states

Let us assume that the stochastic matrix (3) with a single irreducible closed subset of $t \geqslant 2$ states can be written

$$\mathbf{p} = \left(\begin{array}{c|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{Q} & \mathbf{R} \end{array}\right) = \left(\begin{array}{ccc|ccc} p_{11} & \cdots & p_{1t} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{t1} & \cdots & p_{tt} & 0 & \cdots & 0 \\ \hline p_{t+1,1} & \cdots & p_{t+1,t} & p_{t+1,t+1} & \cdots & p_{t+1,s} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{s1} & \cdots & p_{st} & p_{s,t+1} & \cdots & p_{ss} \end{array}\right), \tag{4}$$

where the transition probabilities are $p_{ij} = p_{ij}(\theta)$, some of which may be zero, and where the states $E_1, \ldots, E_t$ form a positively regular chain. It is always possible in a sufficiently large number of transitions to pass from any state $E_{t+1}, \ldots, E_s$ into the subset of states $E_1, \ldots, E_t$; once the system has entered this closed subset no exit from it is possible.

Consider now a realization of the same kind as that for positively regular chains, that is, one starting with a fixed initial state and consisting of a fixed number n of transitions. If we started with one of $E_1, \ldots, E_t$, we should be dealing with the case of the positively regular

chain already investigated; we therefore begin with one of $E_{t+1}, \ldots, E_s$, and obtain a realization S of the $n+1$ states

$$E_i, \quad E_j, \quad \ldots, \quad E_k, \quad E_l,$$

for which the likelihood function is

$$L(S) = \sum_{i,j=1}^{s} n_{ij} \ln p_{ij},$$

where the n_{ij} are frequencies of transition from E_i to E_j, such that $\sum_{i,j=1}^{s} n_{ij} = n$.

It can be verified directly that the form (1) for the $p_{ij}(\theta)$ will satisfy, exactly as in the positively regular case, the sufficiency conditions for the maximum-likelihood estimator T of θ. For the likelihood function is then

$$L(S) = \sum_{i,j=1}^{s} n_{ij} K_{ij} \Lambda_1(\theta) + n\lambda_2(\theta) + \sum_{i,j=1}^{s} n_{ij} \ln \alpha_{ij},$$

and gives the estimator T of θ as

$$\sum_{i,j=1}^{s} n_{ij} K_{ij}/n = -\lambda_2'(T)/\Lambda_1'(T) = g(T),$$

so that $L(S)$ is clearly factorizable as

$$L(S) = n\{g(T)\,\Lambda_1(\theta) + \lambda_2(\theta)\} + \sum_{i,j=1}^{s} n_{ij} \ln \alpha_{ij}. \tag{5}$$

It follows that T is sufficient; moreover, as in Rao (1952, § 4*a*.3) it can be shown that

$$\sum_{i,j} n_{ij} K_{ij}/n = -\lambda_2'(T)/\Lambda_1'(T)$$

is an unbiased estimator of $g(\theta) = -\lambda_2'(\theta)/\Lambda_1'(\theta)$, so that

$$\mathscr{E}\left(\sum_{i,j} n_{ij} K_{ij}/n\right) = g(\theta).$$

A point of minor interest which follows from the structure of the matrix (4) is that the number of distinct forms for the $p_{ij}(\theta)$ in any one of the rows $1, \ldots, t$ is $r \leqslant t$; it follows that there can only be $r \leqslant t$ distinct forms for the p_{ij} in the remaining rows $t+1, \ldots, s$ of the matrix.

3. Chains with a single absorbing state

A typical stochastic matrix with an absorbing state E_1 can be written as

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix}, \tag{6}$$

where the transition probabilities $p_{ij} = p_{ij}(\theta)$, some of which may be zero, are such that if a realization starting with one of $E_2, \ldots, E_s$ is allowed to continue, E_1 must eventually be reached, the total number of transitions in the process being x, a random variable whose distribution depends on θ. We refer to this type of realization as sequential.

Let such a realization result in the sequence S' of states

$$E_i, \quad E_j, \quad \ldots, \quad E_k, \quad E_1;$$

the likelihood function is

$$L(S') = \sum_{i,j} n_{ij} \ln p_{ij} \quad (i = 2, \ldots, s;\ j = 1, \ldots, s),$$

where the n_{ij} are transition frequencies from E_i to E_j, such that $\sum_{i,j} n_{ij} = x$. If the transition probabilities $p_{ij}(\theta)$ except for $p_{11} = 1$ are of the form (1), the likelihood function

$$L(S') = \sum_{i,j} n_{ij} K_{ij} \Lambda_1(\theta) + x\lambda_2(\theta) + \sum_{i,j} n_{ij} \ln \alpha_{ij} \tag{7}$$

will give the maximum-likelihood estimator T_x of θ as

$$\sum_{i,j} n_{ij} K_{ij} / x = -\lambda_2'(T_x)/\Lambda_1'(T_x) = g(T_x). \tag{8}$$

Since x is itself a random variable, the likelihood function $L(S')$ of (7) is not factorizable, as was $L(S)$ of equation (5), nor is T_x a sufficient estimator of θ.

A lemma due to E. Fay, quoted by Lehmann (1950), and proved under somewhat different conditions by Blackwell (1947), enables us to deduce that (T_x, x) is a sufficient statistic* for θ. The lemma states that:

If for each value m of x, T_m is a sufficient statistic for θ in the sample of fixed size $m = \sum_{i,j} n_{ij}$, then (T_x, x) will be a sufficient statistic of θ in the sequential case.

For in the sequence S', starting with E_i, the conditional probability of the sequence, given that $T_m = t$ and $x = m$, is

$$\begin{aligned} \Pr\{S' \mid T_m = t, x = m\} &= \frac{\Pr\{S', T_m = t, x = m\}}{\sum_{S'} \Pr\{S', T_m = t, x = m\}} \\ &= \frac{\exp\{mg(t)\,\Lambda_1(\theta) + m\lambda_2(\theta) + \sum_{i,j} n_{ij} \ln \alpha_{ij}\}}{\sum_{S'} \exp\{mg(t)\,\Lambda_1(\theta) + m\lambda_2(\theta) + \sum_{i,j} n_{ij} \ln \alpha_{ij}\}} \\ &= \frac{\exp\{\sum_{i,j} n_{ij} \ln \alpha_{ij}\}}{\sum_{S'} \exp\{\sum_{i,j} n_{ij} \ln \alpha_{ij}\}}, \end{aligned} \tag{9}$$

where $\sum_{S'}$ indicates summation over all possible realizations S' beginning with E_i, consisting of m transitions, and for which $T_m = t$. Since the probability (9) is independent of θ, (T_x, x) is a sufficient statistic for θ in the sequential case.

In general, however, the maximum-likelihood estimator (8) of $g(\theta)$ will no longer be unbiased. We shall not consider general methods of finding unbiased sufficient estimators of $g(\theta)$, but restrict our discussion to the simple case where $\sum_{i,j} n_{ij} K_{ij}$ and x are linearly related. We show that in some of these cases, unbiased sufficient estimators reduce to those found by Girshick *et al.* (1946) for certain sequential binomial problems.

4. Regular chains for which $\sum_{i,j} n_{ij} K_{ij}$ and x are linearly related

Particularly simple results follow when there is a linear relation between the statistic $\sum_{i,j} n_{ij} K_{ij}$ and the number of transitions x in a sequential realization. Let the relation be

$$\sum_{i,j} n_{ij} K_{ij} = Ax + B, \tag{10}$$

* Although apparently not in universal use, the term is self-explanatory; we follow Lehmann and others in employing it.

where A and B are constants independent of x; we see that the factorizability condition for the likelihood function (7) will now hold, since

$$L(S') = x\{A\Lambda_1(\theta)+\lambda_2(\theta)\}+B\Lambda_1(\theta)+\sum_{i,j} n_{ij}\ln\alpha_{ij}, \tag{11}$$

and the maximum-likelihood estimator T_x of θ is given by

$$-\{B\Lambda_1'(T_x)\}/\{A\Lambda_1'(T_x)+\lambda_2'(T_x)\} = x.$$

It is easily shown that

$$\mathscr{E}(x) = -\{B\Lambda_1'(\theta)\}/\{A\Lambda_1'(\theta)+\lambda_2'(\theta)\} = h(\theta),$$

so that x is an unbiased estimate of $h(\theta)$. We consider some simple examples as illustrations.

The simplest stochastic matrix with an absorbing state is that for the two-state case,

$$\mathbf{p} = \begin{pmatrix} 1 & 0 \\ \exp\{K_{21}\ln\theta(1-\theta)^{-1}+\ln(1-\theta)\}, & \exp\{K_{22}\ln\theta(1-\theta)^{-1}+\ln(1-\theta)\} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1-\theta & \theta \end{pmatrix}, \tag{12}$$

where $K_{21} = 0$, $K_{22} = 1$. For a realization in x transitions, the system starting from E_2 must move from E_2 to E_1 only in the last transition. The relation (10), which is

$$\sum_{i,j} n_{ij}K_{ij} = (x-1)K_{22}+K_{21} = x-1,$$

is clearly satisfied, and the maximum-likelihood estimator $T_x = (x-1)/x$ of θ sufficient.

The three-state case, with the stochastic matrix of the form

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha_{21}\exp K_{21}\Lambda_1+\lambda_2 & \alpha_{22}\exp K_{22}\Lambda_1+\lambda_2 & \alpha_{23}\exp K_{23}\Lambda_1+\lambda_2 \\ \alpha_{31}\exp K_{31}\Lambda_1+\lambda_2 & \alpha_{32}\exp K_{32}\Lambda_1+\lambda_2 & \alpha_{33}\exp K_{33}\Lambda_1+\lambda_2 \end{pmatrix},$$

is known to have the same two or three distinct values of the exponents K_{ij} in each row, with some of the α_{ij} possibly zero. It is found, on considering realizations of the chain starting from a fixed state E_i $(i=2,3)$, that the linear relation (10) will hold only for stochastic matrices

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & 0 \\ 1-\theta & \alpha\theta & (1-\alpha)\theta \\ 1-\theta & \beta\theta & (1-\beta)\theta \end{pmatrix} \quad (0\leqslant\alpha,\beta\leqslant 1), \tag{13}$$

when it takes the form $\sum_{i,j} n_{ij}K_{ij} = x-1$, or

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & 0 \\ 1-\theta & \theta & 0 \\ 0 & 1-\theta & \theta \end{pmatrix}, \tag{14}$$

when it is $\sum_{i,j} n_{ij}K_{ij} = x-r+1$, depending on the initial state E_r $(r=2,3)$.

The solutions (13) and (14) are exhaustive for the three-state case, but it is difficult to treat the general case of the chain with s states in a systematic manner. Nevertheless, it can easily be verified that linear relations (10) hold for stochastic matrices

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1-\theta & \alpha_{22}\theta & \dots & \alpha_{2s}\theta \\ \dots & \dots & \dots & \dots \\ 1-\theta & \alpha_{s2}\theta & \dots & \alpha_{ss}\theta \end{pmatrix} \quad \left(0\leqslant\alpha_{ij}\leqslant 1;\ \sum_{j=2}^{s}\alpha_{ij}=1,\ i=2,\dots,s\right), \tag{15}$$

which are a generalization of (13), or

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1-\theta & \theta & 0 & \dots & 0 \\ 0 & 1-\theta & \theta & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1-\theta & \theta \end{pmatrix}, \tag{16}$$

a generalization of (14), or also stochastic matrices of the mixed form

$$\mathbf{p} = \left(\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ 1-\theta & \alpha_{22}\theta & \dots & \alpha_{2k}\theta & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 1-\theta & \alpha_{k2}\theta & \dots & \alpha_{kk}\theta & 0 & \dots & \dots & \dots \\ \hline 0 & \dots & \dots & 1-\theta & \theta & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 1-\theta & \theta & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots & \dots & 1-\theta & \theta \end{array}\right)$$

$$\left(k<s;\ 0\leqslant\alpha_{ij}\leqslant 1;\ \sum_{j=2}^{k}\alpha_{ij}=1,\ i=2,\dots,k\right). \tag{17}$$

If we assume that the initial state in a realization is the last state E_s, the stochastic matrices (16) of which (12) and (14) are particular cases when s is 2 and 3 respectively, correspond to part of a curtailed single sampling scheme of Girshick *et al.* (1946, § 3B). In the matrix (16), $1-\theta$ is the probability of finding a defective in the sample, and sampling continues until $s-1$ defectives are observed. The likelihood function of a realization S' is clearly

$$L(S') = (s-1)\ln(1-\theta)+(x-s+1)\ln\theta,$$

and the maximum-likelihood estimator of θ

$$T_x = 1-(s-1)/x.$$

This, however, is biased; for, from the probability of a realization

$$\Pr(S') = \binom{x-1}{s-2}(1-\theta)^{s-1}\theta^{x-s+1} \quad (x=s-1,\dots),$$

we can obtain the moment-generating function of x

$$M(t) = \mathscr{E}(e^{tx}) = (1-\theta)^{s-1}(e^{-t}-\theta)^{-s+1},$$

and from it, by first integrating and then substituting $t=0$, we find

$$\left\{\int M(t)\,dt\right\}_{t=0} = \mathscr{E}(x^{-1}) = -(1-\theta)^{s-1}\left\{\theta^{-s+1}\ln(1-\theta)+\sum_{r=2}^{s-1}\frac{(-1)^r}{(r-1)}\theta^{-s+r}(1-\theta)^{-r+1}\right\},$$

so that the expectation of T_x is not θ.

Girshick *et al.* (1946) give the unbiased sufficient estimator of θ for values of $s>2$ as

$$\hat{\theta} = 1-(s-2)/(x-1).$$

If we consider $$\left\{\int e^{-t}M(t)\,dt\right\}_{t=0} = \mathscr{E}((x-1)^{-1}) = (1-\theta)/(s-2),$$

we can verify directly that the expectation of the estimator $\hat{\theta}$ is θ. The case of $s = 2$ is somewhat special; Girshick *et al.* have shown that here, $\hat{\theta}$ can only be 0 or 1.

The previous method for finding unbiased sufficient estimators of θ does not apply to matrices of the forms (15) or (17); for these, the necessary condition that the probability of different paths in a realization of x transitions be the same does not hold.

5. Non-regular finite, and infinite chains satisfying the linear relation

The condition (10) that the estimator of θ in regular chains with an absorbing state be sufficient, may apply equally to some non-regular or infinite chains with absorbing states, for which realizations start from some fixed state E_i and continue until an absorbing state is reached in x transitions. Suppose that, associated with all non-absorbing states E_i, there are only two distinct values of the transition probabilities p_{ij}: θ and $1-\theta$, the remaining values being zero. Then, grouping together the n_{ij} for which the probabilities are θ, and also those for which they are $1-\theta$, so that

$$\sum_{\theta} n_{ij} = n_1, \quad \sum_{1-\theta} n_{ij} = n_2, \quad n_1 + n_2 = x,$$

the likelihood function of the realization is simply

$$L = n_1 \ln \theta + n_2 \ln (1-\theta), \tag{18}$$

and the maximum-likelihood estimator of θ

$$T_x = n_1/(n_1+n_2) = n_1/x.$$

Providing the linear relation (10), which reduces to

$$n_1 = Ax + B, \tag{19}$$

holds, the likelihood function (18) is factorizable as in (11), and the estimator T_x is sufficient.

A non-regular finite chain with s states which satisfies these conditions is the random walk with two absorbing end states, with the stochastic matrix

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \dots \\ 1-\theta & 0 & \theta & \dots & \dots & \dots & \dots \\ 0 & 1-\theta & 0 & \theta & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 1-\theta & 0 & \theta \\ \dots & \dots & \dots & \dots & \dots & 0 & 1 \end{pmatrix}.$$

For a realization starting with E_i $(i = 2, \dots, s-1)$, and ending with E_1 after x transitions, the linear relation (19) is

$$n_1 = \tfrac{1}{2}(x - i + 1), \tag{20}$$

and the estimator of θ,

$$T_x = \tfrac{1}{2} + \tfrac{1}{2}(1-i)/x, \tag{21}$$

will be sufficient. A similar result will hold if the process ends with E_s.

The sufficiency conditions also hold for certain infinite random walks with one absorbing state; for the chain with the infinite stochastic matrix

$$\mathbf{p} = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots \\ 1-\theta & 0 & \theta & \dots & \dots \\ 0 & 1-\theta & 0 & \theta & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad (\theta < \tfrac{1}{2}), \tag{22}$$

in which a realization starts with E_i $(i=2,\ldots)$ and ends with E_1 after x steps, the relation (20) between n_1 and x holds again, and ensures the sufficiency of the estimator (21) of θ. Another infinite random walk is represented by

$$\mathbf{p}=\begin{pmatrix} 1 & 0 & 0 & \ldots & \ldots & \ldots \\ 1-\theta & 0 & 0 & \theta & \ldots & \ldots \\ 0 & 1-\theta & 0 & 0 & \theta & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} \quad (\theta<\tfrac{1}{3}),$$

which for a similar realization will lead to the linear relation

$$n_1=\tfrac{1}{3}(x-i+1),$$

which again ensures the sufficiency of the estimator

$$T_x=\tfrac{1}{3}+(1-i)/(3x)$$

of θ.

Several illustrative examples of the same kind can be contructed; for each, the method of Girshick *et al.* would lead to an unbiased estimator of θ which would be a function of the sufficient estimator T_x. We restrict ourselves to a single example mentioned by Moran, which reduces to a case already described.

Moran (1953) considers the infinite random walk with an absorbing barrier for which the states are $E_{-s},\ldots,E_0,\ldots,E_n,\ldots$, where E_{-s} is the absorbing state. The stochastic matrix for this chain is (22); the process starts from E_0, and proceeds until either E_{-s} is reached in $x\leqslant n$ steps, or any of the other available states E_i $(i=-(s-1),\ldots,0,\ldots,n)$ is reached in n steps. The likelihood functions in these two cases are respectively given as

$$L_1=\tfrac{1}{2}(x-s)\ln\theta+\tfrac{1}{2}(x+s)\ln(1-\theta),\quad L_2=\tfrac{1}{2}(n-s+i)\ln\theta+\tfrac{1}{2}(n+s-i)\ln(1-\theta).$$

The maximum-likelihood estimator T_x of θ is found to be

$$T_x\equiv\{T_x^{(1)}=(x-s)/(2x)\quad\text{or}\quad T_x^{(2)}=(n-s+i)/(2n)\},$$

where, since

$$T_x^{(2)}>\tfrac{1}{2}-s/(2n),\quad T_x^{(1)}=\tfrac{1}{2}-s/(2x)\leqslant\tfrac{1}{2}-s/(2n)<T_x^{(2)},$$

the estimator $T_x\equiv\{T_x^{(1)}\text{ or }T_x^{(2)}\}$ will itself specify the appropriate likelihood function, without the need to state whether absorption has occurred or not. We verify that the likelihood functions in the two cases can be written as

$$L_1=\tfrac{1}{2}s(1-2T_x^{(1)})^{-1}\ln\theta(1-\theta)-\tfrac{1}{2}s\ln\theta(1-\theta)^{-1},\quad L_2=nT_x^{(2)}\ln\theta(1-\theta)^{-1}+n\ln(1-\theta),$$

where the first is effectively the form (18), and the second the ordinary binomial case. The estimator T_x of θ is sufficient, but $T_x^{(1)}$ is not unbiased.

The method of Girshick *et al.* (1946) leads to the sufficient unbiased estimator $T\equiv\{T^{(1)}\text{ or }T^{(2)}\}$ of θ, where these are

$$T^{(1)}=1-\frac{(s-1)(x+s)}{2s(x-1)},\quad T^{(2)}=\frac{\binom{n-1}{\frac{1}{2}(n-s+i)}-\binom{n-1}{\frac{1}{2}(n-s-i)}}{\binom{n}{\frac{1}{2}(n-s+i)}-\binom{n}{\frac{1}{2}(n-s-i)}}.$$

From the expressions

$$T^{(1)}=1-\frac{(s-1)\{(1-2T_x^{(1)})^{-1}+1\}}{2\{s(1-2T_x^{(1)})^{-1}-1\}},\quad T^{(2)}=\frac{\binom{n-1}{nT_x^{(2)}}-\binom{n-1}{n(1-T_x^{(2)})-s}}{\binom{n}{nT_x^{(2)}}-\binom{n}{n(1-T_x^{(2)})-s}},$$

it is clear that the unbiased estimator is a function of the maximum-likelihood estimator T_x.

I wish to thank the unknown referee of an earlier paper (Gani, 1955) for suggestions which have been incorporated in § 5, and also Profs. P. A. Moran and M. S. Bartlett, and Dr H. Ruben for criticisms of an earlier draft of the paper.

REFERENCES

BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge University Press.

BLACKWELL, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* **18**, 105–10.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley.

GANI, J. (1955). Some theorems and sufficiency conditions for the maximum likelihood estimator of an unknown parameter in a simple Markov chain. *Biometrika*, **42**, 342–59.

GIRSHICK, M. A., MOSTELLER, F. & SAVAGE, L. J. (1946). Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statist.* **17**, 13–23.

LEHMANN, E. L. (1950). *Notes on the Theory of Estimation.* Lectures delivered at the University of California.

MORAN, P. A. P. (1953). The estimation of the parameters of a birth and death process. *J. R. Statist. Soc.* B, **15**, 241–5.

RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: John Wiley.

SOME ASYMPTOTIC DISTRIBUTION THEORY FOR MARKOV CHAINS WITH A DENUMERABLE NUMBER OF STATES†

By CYRUS DERMAN

Columbia University, New York

The joint asymptotic distribution is derived for certain functions of the sample realizations of a Markov chain with denumerably many states, from which the joint asymptotic distribution theory of estimates of the transition probabilities is obtained. Application is made to a goodness of fit test.

1. Introduction

Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables which assume only non-negative integer values‡ and which have the property that

$$\Pr\{X_{m+n} = j \mid X_m = i,\ X_{m-1},\ \ldots,\ X_1\}$$
$$= \Pr\{X_{m+n} = j \mid X_m = i\} = \Pr\{X_n = j \mid X_0 = i\}$$

for all integers $m, n \geqslant 0$ and all states i and j for which the conditional probability is defined. The distribution of X_0 will be fixed but arbitrary. Sequences of this type are known as Markov chains with a denumerable number of states and stationary transition probabilities. The fundamentals of the theory were laid down by Kolmogorov (1936). Feller (1950) has also given an exposition of the main results of the theory.

The probability given above is called the nth step transition probability. We shall denote it by $p_{ij}^{(n)}$. When $n = 1$ we shall simply write $p_{ij}^{(1)} = p_{ij}$.

Suppose such a model is assumed and it is of interest to estimate or test hypotheses about the probabilities p_{ij} on the basis of a set of observations $x_1, \ldots, x_n$. We shall develop here some asymptotic distribution theory suitable for this purpose. Results of a similar nature were given by Bartlett (1951). However, he assumed a Markov chain with only a finite number of states. Anderson & Goodman (1956) also considered inference problems for such chains. Our main tool will be a central limit theorem of Doeblin§ (1938) for denumerable Markov chains.

In the remainder of the introduction we shall state, without proof, those results from the general theory which we shall need. The nth step transition probabilities satisfy the relation

$$p_{ij}^{(n)} = \sum_{k=0}^{\infty} p_{ik}^{(n-1)} p_{kj} \quad (i, j = 0, 1, \ldots;\ n = 2, 3, \ldots).$$

More generally

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)} \quad (i, j = 0, 1, \ldots;\ n, m = 1, 2, \ldots).$$

The latter equations are known as the Chapman–Kolmogorov equations.

† Work supported in part by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

‡ Each integer denotes a possible state of the Markov chain.

§ I am indebted to Dr T. E. Harris for calling my attention to the applicability of Doeblin's theorem in connexion with this paper.

An important notion in Markov chain theory is that of the number of transitions necessary to reach a certain state for the first time given initially a particular state. We let

$$f_{ij}^{(n)} = \Pr\{X_{m+n} = j,\ X_{m+\nu} \neq j \text{ for } 1 \leqslant \nu < n \mid X_m = i\} \quad (m, i, j = 0, 1, \ldots;\ n = 1, 2, \ldots);$$

$$f_{ij}^{*} = \sum_{\nu=1}^{\infty} f_{ij}^{(\nu)}, \quad m_{ij} = \sum_{\nu=1}^{\infty} \nu f_{ij}^{(\nu)}, \quad \sigma_{ij} = \sum_{\nu=1}^{\infty} (\nu - m_{ij})^2 f_{ij}^{(\nu)} \quad (m_{ij} < \infty).$$

We call m_{ij} the mean first passage time from state i to state j. If $i = j$, m_{ii} is called the mean recurrence time of state i. States i and j are said to belong to the same class if $f_{ij}^{*} f_{i}^{*} > 0$, i.e. if there exist integers $n(i,j)$ and $n(j,i)$ such that $p_{ij}^{(n(i,j))} > 0$ and $p_{ij}^{(n(j,i))} > 0$. A state i is called recurrent if $f_{ii}^{*} = 1$. This implies that $f_{ji}^{*} = 1$ for all j belonging to the same class as i. If $f_{ii}^{*} < 1$, i is called transient. If i is recurrent and j belongs to the same class as i, then j is also recurrent. Consequently, all states of a class are either all recurrent or all transient. We can then speak of a class as being recurrent or transient. If all states belong to the same class we refer to the Markov chain as being irreducible recurrent (transient). Even though it might not always be explicitly stated, we shall deal throughout only with irreducible chains. If $f_{ii}^{*} = 1$ and $m_{ii} < \infty$, then i is said to be a positive state: if $m_{ii} = \infty$, i is called null. If i is positive (null) and j belongs to the same class, then j is positive (null). Thus the states of a recurrent class will all be positive or all null.

If $f_{ii}^{*} = 1$, $m_{ii} < \infty$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} p_{ji}^{(\nu)} = \frac{1}{m_{ii}}$ for all states j belonging to the same class as i. If $m_{ii} = \infty$, then $\lim_{n\to\infty} p_{ji}^{(n)} = 0$. If all states belong to the same positive class, then

$$\sum_{i=0}^{\infty} \frac{1}{m_{ii}} = 1 \quad \text{and} \quad \sum_{i=0}^{\infty} \frac{1}{m_{ii}} p_{ij} = \frac{1}{m_{jj}} \quad \text{for} \quad j = 0, 1, \ldots.$$

We call a state i periodic with period t if t is the greatest common divisor of all n such that $p_{ii}^{(n)} > 0$. All states in a class have the same period. A chain, if irreducible, is called periodic or aperiodic according as $t > 1$ or $= 1$.

We shall consider the number of times a state is visited in n transitions. Let

$$N_n(i) = \{\text{the number of } v\text{'s such that } X_v = i \text{ for } 1 \leqslant v \leqslant n\}$$

for $i = 0, 1, \ldots;\ n = 1, \ldots$. More generally we define

$$N_n(i_1, \ldots, i_r) = \{\text{the number of } v\text{'s such that } X_v = i_r, \ldots, X_{v-r+1} = i_1 \text{ for } 1 \leqslant v \leqslant n\},$$

where $i_1, \ldots, i_r$ is a finite subset of all states.

2. Some limit theorems

We shall assume throughout that the Markov chain under consideration is irreducible, positive recurrent and aperiodic.

We first state, without proof, a theorem proved by Doeblin (1938). Consider the sequence of random variables $\{Y_n\}$ attached to the Markov chain $\{X_n\}$ in the following way:

$Y_n = x_i$ if $X_n = i$ $(i = 0, 1, \ldots)$, where the x_i's are arbitrary real numbers. Let

$$U_l(i) = Y_{m_l+1} + \ldots + Y_{m_{l+1}},$$

where m_l is the trial in which state i is reached for the lth time. It follows that $\{U_l(i)\}$ $(l = 1, 2, \ldots)$ is a sequence of independent and identically distributed random variables.

Let $g(j) = x_j - EU_i(i)/m_{ii}$ (the dependence on i is only apparent; cf. (2·1)).

THEOREM 1 (Doeblin). *If $EU_i^2(i) < \infty$ and $\sigma_{ii}^2 < \infty$, then the distribution of*

$$\left\{\sum_{m=1}^{n} Y_m - \frac{n}{m_{ii}} EU_i(i)\right\} n^{-\frac{1}{2}}$$

tends as $n \to \infty$ to a Normal distribution with mean zero and variance

$$\sigma^2 = \sum_{j=0}^{\infty} \frac{g^2(j)}{m_{jj}} + 2 \sum_{j=0, j\neq i}^{\infty} \frac{g(j)}{m_{jj}} \sum_{k=0}^{\infty} \frac{m_{ji} + m_{ik} - m_{jk}}{m_{kk}} g(k)$$

provided that the series converge absolutely.

The expression for σ^2 is due to Chung (1953). Using Doeblin's theorem, we now prove

THEOREM 2. *Let $i_1, \ldots, i_r$ be any finite sequence of distinct states belonging to the same positive class and suppose there exists a state i in the class such that $\sigma_{ii}^2 < \infty$, then there is a matrix Σ such that the distribution of*

$$\frac{1}{n^{\frac{1}{2}}}\left\{N_n(i_1) - \frac{n}{m_{i_1 i_1}}, \ldots, N_n(i_r) - \frac{n}{m_{i_r i_r}}\right\}$$

tends as $n \to \infty$ to a multivariate Normal distribution with mean zero and covariance matrix Σ, which can be found from the expression for σ^2 in Doeblin's theorem.

Proof. Chung (1953) has shown that

$$\left.\begin{aligned} EU_i(i) &= m_{ii} \sum_{j=0}^{\infty} \frac{x_j}{m_{jj}}, \\ EU_i^2(i) &= m_{ii} \sum_{j=0}^{\infty} \frac{x_j^2}{m_{jj}} + 2m_{ii} \sum_{\substack{j=0 \\ (j\neq i)}}^{\infty} \frac{x_j}{m_{jj}} \sum_{k=0}^{\infty} \frac{m_{ji} + m_{ik} - m_{jk}}{m_{kk}} x_k, \end{aligned}\right\} \tag{2·1}$$

provided that the series converge absolutely.

Let $x_{i_1} = \lambda_1, \ldots, x_{i_r} = \lambda_r$, $x_j = 0$ for all $j \neq i_1, \ldots, i_r$, where $\lambda_1, \ldots, \lambda_r$ is any arbitrary set of real numbers. It is clear from (2·1) that $EU_i^2(i) < \infty$ and

$$EU_i(i) = m_{ii} \sum_{\alpha=1}^{r} \frac{\lambda_\alpha}{m_{i_\alpha i_\alpha}}. \tag{2·2}$$

Then

$$\begin{aligned} \sum_{m=1}^{n} Y_m - \frac{n}{m_{ii}} EU_i(i) &= \sum_{m=1}^{n} Y_m - n \sum_{\alpha=1}^{r} \frac{\lambda_\alpha}{m_{i_\alpha i_\alpha}} \\ &= \sum_{\alpha=1}^{r} \lambda_\alpha \left(N_n(i_\alpha) - \frac{n}{m_{i_\alpha i_\alpha}}\right). \end{aligned} \tag{2·3}$$

Because of Theorem 1 and the continuity theorem for characteristic functions,

$$\lim_{n\to\infty} E \exp\left\{i \sum_{\alpha=1}^{r} \lambda_\alpha \left(N_n(i_\alpha) - \frac{n}{m_{i_\alpha i_\alpha}}\right) n^{-\frac{1}{2}}\right\} = e^{-1/2Q}, \tag{2·4}$$

where $Q = \sigma^2$ is a positive semi-definite quadratic function of $\lambda_1, \ldots, \lambda_r$. But since the λ's are arbitrary (2·4) indicates the convergence of the distribution of the vector

$$\frac{1}{n^{\frac{1}{2}}}\left\{N_n(i_1) - \frac{n}{m_{i_1 i_1}}, \ldots, N_n(i_r) - \frac{n}{m_{i_r i_r}}\right\} \tag{2·5}$$

to a multivariate Normal distribution. The covariance matrix Σ is determined by Q.

The fact that only one state i such that $\sigma_{ii}^2 < \infty$ is necessary for the truth of Theorem 2 suggests the possibility that the existence of one such state i might imply $\sigma_{jj}^2 < \infty$ for all j in the same class. As pointed out in the introduction, the first recurrence moment is finite

for all or no states of the same class. The following theorem, proved by Chung (1954), is a generalization of this fact and confirms the truth of the above conjecture:

THEOREM 3 (Chung). *The ρ-th moment, $\rho > 0$, of the recurrence time of i is finite, if and only if for every j and k belonging to the same class, the ρ-th moments of the first passage times from j to k and from k to j are finite.*

Let $\{X_n(r)\}$ denote a Markov chain derived from $\{X_n\}$ as follows: $X_n(r) = (i_1, \ldots, i_r)$ when $X_n = i_r, \ldots, X_{n-r+1} = i_1$ for every set of states $i_1, \ldots, i_r$. States having probability zero of occurring will be ignored.

LEMMA 1. *If the ρ-th recurrence moments for the states of $\{X_n\}$ are finite, then the ρ-th recurrence moments for the states of $\{X_n(r)\}$ are finite.*

Proof. Consider any state $(i_1, \ldots, i_r)$ which has positive probability of occurring. Suppose $X_0 = i_1$. Let τ be a random variable which denotes the number of steps until the rth recurrence of i_1. Since τ is the sum of r independent and identically distributed random variables (the number of steps until one recurrence of i_1) each of which has by hypothesis a finite ρth moment, it follows easily that $E\tau^\rho < \infty$. We shall say the event ϵ is associated with τ if $X_\nu(r) = (i_1, \ldots, i_r)$ for some $\nu\ (r-1 \leqslant \nu < \tau)$; i.e. the event ϵ is associated with a random variable τ if during the course of the r recurrences of i the sequence of states $i_1, \ldots, i_r$ is visited in succession. Let $\tau_1, \ldots, \tau_N$ denote N successive such random variables where N is the smallest possible integer such that the event ϵ is associated with two of the random variables. The N random variables are independently distributed. τ_N and one other random variable have for their distributions the conditional distribution of τ given that ϵ is associated with it, and the $N-2$ others have for their distributions the condition distribution of τ given that ϵ is not associated with it. Now, if it can be shown that $E(\tau_1 + \ldots + \tau_N)^\rho < \infty$, it will follow, since at least two occurrences of the state $(i_1, \ldots, i_r)$ take place during the time $\tau_1 + \ldots + \tau_N$, that the ρth recurrence moment of the state $(i_1, \ldots, i_r)$ is finite. To this end let P denote the probability that the event ϵ is associated with τ. Since $(i_1, \ldots, i_r)$ has positive probability of occurring, $P > 0$. If $P = 1$, then $N = 2$ with probability 1, and it is easily seen that $E(\tau_1 + \tau_2)^\rho < \infty$. Suppose $P < 1$. Let $E(\tau^\rho \mid \epsilon)$ and $E(\tau^\rho \mid \bar{\epsilon})$ denote the conditional expectations of τ^ρ given that ϵ is associated and not associated with τ^ρ respectively. Since

$$E\tau^\rho = E(\tau^\rho \mid \epsilon)\, P + E(\tau^\rho \mid \bar{\epsilon})\,(1-P) < \infty$$

it follows that $E(\tau^\rho \mid \epsilon)$ and $E(\tau^\rho \mid \bar{\epsilon})$ are both finite. It is clear that

$$\Pr(N = v) = (v-1)\, P^2 (1-P)^{v-2} \quad (v = 2, \ldots),$$

i.e. N has a Pascal distribution. Therefore

$$\begin{aligned} E(\tau_1 + \ldots + \tau_N)^\rho &= \sum_{v=2}^{\infty} E(\tau_1 + \ldots + \tau_v)^\rho \Pr(N = v) \\ &\leqslant \sum_{v=2}^{\infty} E(v \max(\tau_1, \ldots, \tau_v))^\rho\, (v-1)\, P^2 (1-P)^{v-2} \\ &= \sum_{v=2}^{\infty} v^\rho E(\max(\tau_1, \ldots, \tau_v))^\rho\, (v-1)\, P^2 (1-P)^{v-2} \\ &\leqslant \sum_{v=2}^{\infty} v^{\rho+2} \max(E(\tau^\rho \mid \epsilon),\ E(\tau^\rho \mid \bar{\epsilon}))\, P^2 (1-P)^{v-2} \\ &< \infty. \end{aligned}$$

This proves the lemma.

Now let $i_{11}, \ldots, i_{1r_1}; i_{21}, \ldots, i_{2r_2}; \ldots; i_{k1}, \ldots, i_{kr_k}$ be k sets of states each having $r_1, \ldots, r_k$ members respectively. Let

$$P_l = \prod_{\alpha=1}^{r_l-1} p_{i_{l\alpha} i_{l(\alpha+1)}} \quad (l = 1, \ldots, k).$$

We prove now a slight, but useful, generalization of Theorem 2.

THEOREM 4. *If $i_{11}, \ldots, i_{kr_k}$ belong to the same class and if $\sigma_{ii}^2 < \infty$ for i in the same class, then as $n \to \infty$ the distribution of*

$$\frac{1}{n^{\frac{1}{2}}}\left\{N_n(i_{11}, \ldots, i_{1r_1}) - \frac{nP_1}{m_{i_{11}i_{11}}}, \ldots, N_n(i_{k1}, \ldots, i_{kr_k}) - \frac{nP_k}{m_{i_{k1}i_{k1}}}\right\} \tag{2·6}$$

tends to a multivariate Normal distribution with mean zero and covariance matrix $\mathfrak{A}$ to be determined later.

Proof. Let $r = \max(r_1, \ldots, r_k)$. Let $\lambda_1, \ldots, \lambda_k$ be arbitrary real numbers and associate λ_l with the sequence of states $i_{l1}, i_{l2}, \ldots, i_{lr_l}$ $(l = 1, \ldots, k)$. Let $\{Y_n(r)\}$ be defined from $X_n(r)$ as follows:

$$Y_n(r) = \sum_a \lambda_l \quad \text{if} \quad X_n(r) = (i_1, \ldots, i_{r-r_l}, i_{l1}, \ldots, i_{lr_l}),$$

where $\sum_a$ denotes the sum over all such possible l for a given state of $X_n(r)$; e.g. if j and i, j are two sets of states and λ_1 and λ_2 are associated with j and i, j respectively, then whenever $X_n(2) = (i, j)$, $Y_n(2) = \lambda_1 + \lambda_2$. If for a given state there is no such l, then $Y_n(r) = 0$. Let $(i_1, \ldots, i_r)$ be any state of $\{X_n(r)\}$. Define $U_l(i_1, \ldots, i_r)$ as in Theorem 2. By Lemma 1 the variance of the recurrence time of $(i_1, \ldots, i_r)$ is finite. We have, also, that $EU_l^2(i_1, \ldots, i_r) < \infty$, since $U_l(i_1, \ldots, i_r) < Z \sum_{l=1}^{k} |\lambda_l|$, where Z is the recurrence time of $(i_1, \ldots, i_r)$. Using the fact that $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=0}^{n} p_{ii}^{(\nu)} = \frac{1}{m_{ii}}$ for all i, it is easy to see that the mean recurrence time of any state $(i_1, \ldots, i_r)$, which we denote by $m(i_1, \ldots, i_r)$, is $\dfrac{m_{i_1 i_1}}{p_{i_1 i_2} \cdots p_{i_{r-1} i_r}}$. Thus, since absolute convergence is easily proved, we have from (2·1)

$$\begin{aligned} \frac{EU(i_1, \ldots, i_r)}{m(i_1, \ldots, i_r)} &= \sum_{\alpha=1}^{k} \sum_c \frac{\lambda_\alpha}{m(i_1, \ldots, i_{r-r_\alpha}, i_1, \ldots, i_r)} \\ &= \sum_{\alpha=1}^{k} \sum_c \frac{\lambda_\alpha p_{i_1 i_2} \cdots p_{i_{r-r_\alpha-1} i_{r-r_\alpha}} p_\alpha}{m_{i_1 i_1}} \\ &= \sum_{\alpha=1}^{k} \frac{\lambda_\alpha p_\alpha}{m_{i_{\alpha 1} i_{\alpha 1}}}, \end{aligned} \tag{2·7}$$

where $\sum_c$ denotes the sum over all possible sequences $i_1, \ldots, i_{r-r_\alpha}$. The argument used in proving Theorem 2 now applies again. This proves the theorem.

The cases of interest to us below will be where the r_l's are either 1 or 2. However, we shall maintain the general notation as long as possible.

For statistical applications of Theorem 4, it is necessary to be able to determine $\mathfrak{A}$, the covariance matrix of (2·6). From Formula 4 below it follows that the vector

$$n^{-\frac{1}{2}}\{N_n(i_{11}, \ldots, i_{1r_1}) - EN_n(i_{11}, \ldots, i_{1r_1}), \ldots, N_n(i_{k1}, \ldots, i_{kr_k}) - EN_n(i_{k1}, \ldots, i_{kr_k})\}, \tag{2·8}$$

and (2·6) have the same limiting distribution. Feller (1949) has shown, assuming the

finiteness of a second recurrence moment, that the limit of the variances of (2·8) as $n \to \infty$ is equal to the variances of the limiting distribution of (2·8). If we can show that the same holds for the covariances then it will follow that $\mathfrak{T}$ can be approximated by computing the covariance matrix of (2·8).

The following theorem is for this purpose.

THEOREM 5. *Suppose the sequence of random vectors (X_n, Y_n) converges in distribution to a joint distribution function $F(x, y)$ which has finite second moments. Let (X, Y) be a random vector with joint-distribution function $F(x, y)$. If $\lim_{n\to\infty} EX_n^2 = EX^2$ and $\lim_{n\to\infty} EY_n^2 = EY^2$, then $\lim_{n\to\infty} EX_nY_n = EXY$.*

Proof. Let $Z_n = X_n + Y_n$, $Z = X + Y$. A consequence of the Helly–Bray theorem is that $EZ^2 \leqslant \liminf_{n\to\infty} EZ_n^2$. This implies $E(XY) \leqslant \liminf_{n\to\infty} E(X_nY_n)$. Now let $Z_n' = X_n - Y_n$, $Z' = X - Y$. In the same way we get that $E(XY) \geqslant \limsup_{n\to\infty} E(X_nY_n)$. Therefore $\lim_{n\to\infty} E(X_nY_n) = E(XY)$, proving the theorem.

3. COMPUTATION OF THE COVARIANCE MATRIX

We now commence with the computation of the covariance matrix of (2·8). It will be convenient to prove some asymptotic formulae which will be needed in the computation. In all that follows the second recurrence moment will be assumed finite.

We state without proof

LEMMA 2. *If $a_n > 0$, $\lim_{n\to\infty} b_n = b$, $\lim_{n\to\infty} \dfrac{a_n}{\sum_{\nu=0}^{n} a_\nu} = 0$, then*

$$\lim_{n\to\infty} \sum_{\nu=0}^{n} a_{n-\nu} b_\nu \Big/ \sum_{\nu=0}^{n} a_\nu = b.$$

Formula 1.

$$\sum_{\nu=1}^{\infty} \left(p_{ij}^{(\nu)} - \frac{1}{m_{jj}}\right) = \frac{M_{jj}}{2m_{jj}^2} - \frac{m_{ij}-1}{m_{jj}} \quad (i, j, = 0, 1, \ldots),$$

where

$$M_{jj} = \sum_{\nu=1}^{\infty} \nu(\nu-1) f_{jj}^{(\nu)}.$$

Proof. We have directly

$$\sum_{n=1}^{N} \left(p_{ij}^{(n)} - \frac{1}{m_{jj}}\right) = \sum_{n=1}^{N} \left\{ \sum_{\nu=1}^{n} f_{ij}^{(\nu)} \left(p_{jj}^{(n-\nu)} - \frac{1}{m_{jj}}\right) \frac{1 - \sum_{\nu=1}^{n} f_{ij}^{(\nu)}}{m_{jj}} \right\}$$

$$= \sum_{\nu=1}^{N} f_{ij}^{(\nu)} \sum_{t=0}^{N-\nu} \left(p_{jj}^{(t)} - \frac{1}{m_{jj}}\right) - \sum_{n=1}^{N} \frac{\left(1 - \sum_{\nu=1}^{n} f_{ij}^{(\nu)}\right)}{m_{jj}}. \tag{3·1}$$

Feller (1949) has shown that

$$\sum_{t=0}^{\infty} \left(p_{jj}^{(t)} - \frac{1}{m_{jj}}\right) = \frac{M_{jj}}{2m_{jj}^2}.$$

We also have that

$$\sum_{n=1}^{\infty} \left(1 - \sum_{\nu=1}^{n} f_{ij}^{(\nu)}\right) = m_{ij} - 1.$$

Hence using Lemma 2 we get Formula 1.

LEMMA 3.

$$\sum_{n=1}^{N} n\left(p_{ji}^{(n)} - \frac{1}{m_{ii}}\right) = o(N) \quad (i, j = 0, 1, \ldots).$$

Proof. As in proving Formula 1 we have

$$\sum_{n=1}^{N} n\left(p_{ji}^{(n)} - \frac{1}{m_{ii}}\right) = \sum_{n=1}^{N} n\left\{\sum_{\nu=1}^{n} f_{ji}^{(\nu)}\left(p_{ii}^{(n-\nu)} - \frac{1}{m_{ii}}\right) - \frac{\left(1 - \sum_{\nu=1}^{n} f_{ji}^{(\nu)}\right)}{m_{ii}}\right\}$$

$$= \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{n=\nu}^{N} n\left(p_{ii}^{(n-\nu)} - \frac{1}{m_{ii}}\right) - \sum_{n=1}^{N} n\,\frac{\left(1 - \sum_{\nu=1}^{n} f_{ji}^{(\nu)}\right)}{m_{ii}}$$

$$= \sum_{\nu=1}^{N} \nu f_{ji}^{(\nu)} \sum_{t=0}^{N-\nu}\left(p_{ii}^{(t)} - \frac{1}{m_{ii}}\right) + \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{t=0}^{N-\nu} t\left(p_{ii}^{(t)} - \frac{1}{m_{ii}}\right) - \sum_{n=1}^{N} n\,\frac{\left(1 - \sum_{\nu=1}^{n} f_{ji}^{(\nu)}\right)}{m_{ii}}. \tag{3·2}$$

By Lemma 2 and Feller's result the first term on the right of (3·2) is bounded. Now

$$\frac{1}{N}\left|\sum_{\nu=1}^{N} f_{ji}^{(\nu)} \sum_{t=0}^{N-\nu} t\left(p_{ii}^{(t)} - \frac{1}{m_{ii}}\right)\right| \leqslant \sum_{\nu=1}^{N} f_{ji}^{(\nu)} \frac{1}{N-\nu} \sum_{t=0}^{N-\nu} t\left|p_{ii}^{(t)} - \frac{1}{m_{ii}}\right|. \tag{3·3}$$

Feller (1949) has shown that $\sum_{t=0}^{\infty}\left|p_{ii}^{(t)} - \frac{1}{m_{ii}}\right| < \infty$. Hence $\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} t\left|p_{ii}^{(t)} - \frac{1}{m_{ii}}\right| = 0$.

Using Lemma 2 we then see that the second term on the right of (3·2) is $o(N)$.

A similar argument, making use of the fact that $\sum_{n=1}^{\infty}\left(1 - \sum_{\nu=1}^{n} f_{ji}^{(\nu)}\right) < \infty$, applies to the third term of (3·2). Hence Lemma 3 is proved.

Formula 2.

$$\sum_{\nu=1}^{n-r} (n-\nu-k)\, p_{ii}^{(\nu)} = n\left(\frac{M_{ii} - 2(m_{ii}-1)\, m_{ii}}{2m_{ii}^2}\right) + \frac{(n-k)\,(n-k-1)}{2m_{ii}} + o(n)$$

for $l, i = 0, 1, \ldots$ and any integers $r, k > 0$.

Proof. We have

$$\sum_{\nu=1}^{n-r} p_{ii}^{(\nu)}(n-\nu-k) = \sum_{\nu=1}^{n-r} (n-\nu-k)\left(p_{ii}^{(\nu)} - \frac{1}{m_{ii}}\right) + \sum_{\nu=1}^{n-r} \frac{n-\nu-k}{m_{ii}}$$

$$= n\sum_{\nu=1}^{n-r}\left(p_{ii}^{(\nu)} - \frac{1}{m_{ii}}\right) - \sum_{\nu=1}^{n-r} (\nu+k)\left(p_{ii}^{(\nu)} - \frac{1}{m_{ii}}\right) + \frac{1}{m_{ii}}\sum_{\nu=1}^{n-r} (n-\nu-k)$$

$$= n\left(\frac{M_{ii} - 2(m_{ii}-1)\, m_{ii}}{2m_{ii}^2}\right) + \frac{(n-k-1)\,(n-k)}{2m_{ii}} + o(n), \tag{3·4}$$

by virtue of Formula 1 and Lemma 3.

Formula 3.

$$\sum_{\nu=1}^{n-r} p_{mi}^{(\nu)} \sum_{t=1}^{n-k-\nu} p_{ij}^{(t)} = \frac{(n-k)\,(n-k-1)}{2m_{ii}m_{jj}} + n\left\{\frac{M_{jj} - 2(m_{ij}-1)\, m_{jj}}{2m_{jj}^2 m_{ii}} + \frac{M_{ii} - 2(m_{mi}-1)\, m_{ii}}{2m_{ii}^2 m_{jj}}\right\} + o(n)$$

for $m, i, j = 0, 1, \ldots$ and all integers $r, k > 0$.

Proof. We have

$$\sum_{\nu=1}^{n-r} p_{mi}^{(\nu)} \sum_{t=1}^{n-k-\nu} p_{ij}^{(t)} = \sum_{\nu=1}^{n-r} p_{mi}^{(\nu)} \sum_{t=1}^{n-k-\nu} \left(p_{ij}^{(t)} - \frac{1}{m_{jj}}\right) + \sum_{\nu=1}^{n-r} (n-k-\nu)\frac{p_{mi}^{(\nu)}}{m_{jj}}. \tag{3·5}$$

Using the fact that $\sum_{\nu=1}^{n} p_{mi}^{(\nu)} = \frac{n}{m_{ii}} + o(n)$, Formula 1, and Lemma 2 on the first term on the right-hand side of (3·5) and Formula 2 on the second term we get that (3·5) yields Formula 3.

We are now in a position to calculate the moments of $N_n(i_1, \ldots, i_r)$.

Formula 4.

$$EN_n(i_1, \ldots, i_r) = P\left\{\frac{n-r+2}{m_{i_1 i_1}} + \frac{M_{i_1 i_1} - 2m_{li_1} m_{i_1 i_1}}{2m_{i_1 i_1}^2}\right\} + o(1),$$

where $P = \prod_{\alpha=1}^{r-1} p_{i_\alpha i_{\alpha+1}}$ and l is the initial state.

Proof. Let $Y_n = 1$ if $X_n = i_r$, $X_{n-1} = i_{r-1}, \ldots, X_{n-r+1} = i_1$; $Y_n = 0$ otherwise. Since $X_0 = l$ we have

$$EN_n(i_1, \ldots, i_r) = E\sum_{\nu=1}^{n} Y_\nu = P\sum_{\nu=1}^{n-r+1} p_{li}^{(\nu)}. \tag{3·6}$$

The result follows by using Formula 1.

Formula 5.

$$\operatorname{var} N_n(i_1, \ldots, i_r) = nP\left\{\frac{1}{m_{i_1 i_1}} + P\left(\frac{1 - 2m_{i_r i_1}}{m_{i_1 i_1}^2} + \frac{M_{i_1 i_1}}{m_{i_1 i_1}^3}\right)\right\} + o(n)$$

for $i_r \neq i_1, \ldots, i_{r-1}$.

Proof. We have

$$EN_n^2(i_1, \ldots, i_r) = E\left\{\sum_{\nu=1}^{n} Y_\nu\right\}^2 = E\left\{\sum_{\nu=r}^{n} Y_\nu^2 + 2\sum_{\nu=r}^{n-1} Y_\nu \sum_{t=\nu+1}^{n} Y_t\right\}$$

$$= EN_n(i_1, \ldots, i_r) + 2P^2 \sum_{\nu=1}^{n-r} p_{li}^{(\nu)} \sum_{t=1}^{n-\nu-r+1} p_{i_r i_1}^{(t)}. \tag{3·7}$$

Application of Formulae 3 and 4, together with the fact that

$$\operatorname{var} N_n(i_1, \ldots, i_r) = EN_n^2(i_1, \ldots, i_r) - (EN_n(i_1, \ldots, i_r))^2$$

proves the result.

Formula 6.

$$\operatorname{cov}(N_n(i), N_n(j)) = n\left\{\frac{1}{m_{ii} m_{jj}} - \frac{(m_{kj} + m_{ji})}{m_{ii} m_{jj}} + \frac{1}{2}\left(\frac{M_{ii}}{m_{ii}^2 m_{jj}} + \frac{M_{jj}}{m_{jj}^2 m_{ii}}\right)\right\} + o(n).$$

Proof. The proof is similar to those of Formulae 4 and 5.

Formula 7.

$$\operatorname{var} N_n(i, i) = \frac{np_{ii}}{m_{ii}}\left\{1 + p_{ii}\left(\frac{1}{m_{ii}} + \frac{M_{ii}}{m_{ii}^2}\right)\right\} + o(n).$$

Proof. The proof is as in the proof of Formula 5 allowing for the fact that the lower limit for t in (3·7) is 0 instead of 1.

Applying the well-known techniques of asymptotic distribution theory we get easily

$$\operatorname{var}\sqrt{n}\frac{N_n(i,j)}{N_n(i)} \sim \frac{m_{ii}^2}{n}\operatorname{var} N_n(i,j) + p_{ij}^2 \frac{m_{ii}^2}{n}\operatorname{var} N_n(i) - 2p_{ij}\frac{m_{ii}^2}{n}\operatorname{cov}(N_n(i), N_n(i,j)),$$

$$\operatorname{cov}\left(\sqrt{n}\frac{N_n(i,j)}{N_n(i)}, \sqrt{n}\frac{N_n(k,l)}{N_n(k)}\right) \sim \frac{m_{ii} m_{kk}}{n}\{\operatorname{cov}(N_n(i,j), N_n(k,l)) - p_{kl}\operatorname{cov}(N_n(i,j), N_n(k))$$

$$- p_{ij}\operatorname{cov}(N_n(i), N_n(k,l)) + p_{kl} p_{ij}\operatorname{cov}(N_n(i), N_n(k))\}. \tag{3·8}$$

Using the above formula and (3·9) and after carrying out the computations necessary to evaluate

$$\operatorname{cov}(N_n(i,j), N_n(k,l)) \quad \text{and} \quad \operatorname{cov}(N_n(i), N_n(j,k))$$

for the various special cases we get the following results:

$$\operatorname{var}\sqrt{n}\frac{N_n(i,j)}{N_n(i)} \sim m_{ii}p_{ij}(1-p_{ij}), \quad \operatorname{cov}\left(\sqrt{n}\frac{N_n(i,j)}{N_n(i)}, \sqrt{n}\frac{N_n(k,l)}{N_n(k)}\right) \sim \begin{cases} -m_{ii}p_{ij}p_{il} & \text{if } i=k, \\ 0 & \text{if } i \neq k. \end{cases} \tag{3·9}$$

The analogy of (3·9) with the second moments of a multinomial distribution, with m_{ii} playing the role of the number of trials starting from i, is clear.

4. Application to a goodness of fit test

Consider the problem of testing the hypothesis that the matrix of transition probabilities is a given matrix p_{ij} on the basis of $X_1, \ldots, X_n$, where X_i $(i = 1, \ldots, n)$ are observed states of the Markov chain in the first n steps. It is clear that a finite number of observations will not provide estimates of all the transition probabilities. This suggests choosing a finite subset of the transition probabilities independently of the data and testing the hypothesis that they are the true transition probabilities, i.e. testing, only partially, the original hypothesis. A rejection of the partial hypothesis would, of course, imply a rejection of the original hypothesis. The question as to how to select the subset of transition probabilities will not be considered. Thus, as far as this paper is concerned, the selection is arbitrary.

Let

$$\hat{p}_{ij} = \frac{N_n(i,j)}{N_n(i)}, \quad Z_{ij} = \sqrt{n}\frac{(\hat{p}_{ij}-p_{ij})}{\sqrt{(m_{ii}p_{ij})}}, \quad R_i(C_i) = 1 - \sum_{j \in C_i} p_{ij},$$

$$\hat{R}_i(C_i) = 1 - \sum_{j \in C_i} \hat{p}_{ij} \quad \text{and} \quad Z_i(C_i) = \frac{\sqrt{n}\,(\hat{R}_i(C_i) - R_i(C_i))}{\sqrt{\{m_{ii}R_i(C_i)\}}},$$

where C_i denotes a finite class of states depending on i. It can be shown using (3·9) that

$$\left.\begin{aligned} \operatorname{var}(Z_{ij}) &\sim (1-p_{ij}), \quad \operatorname{cov}(Z_{ij}, Z_{il}) \sim -\sqrt{(p_{ij}p_{il})} \quad (l \neq j), \\ \operatorname{cov}(Z_{ij}, Z_{kl}) &\sim 0 \quad (i \neq k), \\ \operatorname{cov}(Z_{ij}, Z_i(C_i)) &\sim -\sqrt{\{p_{ij}R_i(C_i)\}}, \\ \operatorname{cov}(Z_{ij}\ Z_k(C_k)) &\sim 0 \quad (k \neq i), \\ \operatorname{var}(Z_i(C_i)) &\sim 1 - R_i(C_i), \quad \operatorname{cov}(Z_i(C_i), Z_k(C_k)) \sim 0 \quad (i \neq k). \end{aligned}\right\} \tag{4·1}$$

It follows from well-known asymptotic distribution theory using Theorem 4, that, if $\{p_{ij}\}$ is the true matrix of transition probabilities, the limiting distribution of any finite set of Z_{ij}'s and $Z_i(C_i)$'s is multivariate Normal with zero means and covariance matrix given by (4·1). Let Z'_{ij} $(j \in C_i)$, $Z'_i(C_i)$ $(i = 1, \ldots, k)$ be random variables with a multivariate normal distribution with zero means and covariance matrix given by (4·1). Then it follows (see Cramér, 1946, p. 419) that

$$\chi'^2 = \sum_{i=1}^{k}\left\{\sum_{j \in C_i} Z'^2_{ij} + Z'^2_i(C_i)\right\}$$

has a χ^2 distribution with $\sum_{i=1}^{k} d_i$ degrees of freedom where d_i denotes the number of states

in C_i. Since the corresponding Z_{ij}'s and $Z_i(C_i)$'s have such a limiting distribution, it also follows (see Cramér, 1946, p. 314) that the limiting distribution of

$$\chi^2 = \sum_{i=1}^{k} \left\{ \sum_{j \in C_i} Z_{ij}^2 + Z_i^2(C_i) \right\}$$

is that of χ'^2. Thus the χ^2 statistic supplies a χ^2 test for the goodness of fit of a given finite set of transition probabilities. The m_{ii}'s in χ^2 can, in principle, be derived from $\{p_{ij}\}$. In most cases this will not be feasible. The limiting distribution of χ^2 will not be changed if the m_{ii}'s are replaced by their estimates obtained from the recurrence times.

I wish to thank Profs. T. W. Anderson and D. A. Darling for several helpful conversations held while this work was in progress.

REFERENCES

ANDERSON, T. W. & GOODMAN, L. (1956). Statistical inference in Markov chains. To appear in *Ann. Math. Statist.*

BARTLETT, M. S. (1951). The frequency goodness of fit test for probability chains. *Proc. Camb. Phil. Soc.* **47**, 86–95.

CHUNG, K. L. (1953). Contributions to the theory of Markov chains. I. *J. Res. Nat. Bur. Stand.* **50**, 203–8.

CHUNG, K. L. (1954). Contributions to the theory of Markov chains. II. *Trans. Amer. Math. Soc.* **76**, 397–419.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

DOEBLIN, W. (1938). Sur deux problèmes de M. Kolmogoroff concernant les chaines dénombrables. *Bull. Soc. math. Fr.* **66**, 210–20.

FELLER, W. (1949). Fluctuation theory of recurrent events. *Trans. Amer. Math. Soc.* **67**, 98–119.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

KOLMOGOROV, A. N. (1936). Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. *Mathematiceskii Sbornik*, N.S., **1**, 607–10.

A GENERAL METHOD FOR APPROXIMATING TO THE DISTRIBUTION OF LIKELIHOOD RATIO CRITERIA

By D. N. LAWLEY

University of Edinburgh

1. Introduction

In the theory of testing statistical hypotheses it is well known that $-2\log\lambda$, where λ is the likelihood criterion of Neyman & Pearson (1928), is distributed for large samples approximately as χ^2. This was proved by Wilks (1938). Box (1949) has shown that for a certain class of such criteria it is possible to improve on this approximation in various ways, so that in cases where the exact distribution of the criterion is unknown good approximations based on a knowledge of the moments of the criterion can be developed even for samples of moderate size. The simplest improvement consists in multiplying $-2\log\lambda$ by a scale factor which results in a statistic having the same moments as χ^2 ignoring quantities of order n^{-2}, where n is the size of the sample. This scaling device was first used by Bartlett (1937; see also his recent note, 1954). The object of this paper is to show that for any likelihood function satisfying certain very general conditions an improved χ^2 test of this type is, in theory, possible.

2. An expansion for $L^{(k)}$

Let us suppose that the likelihood function, whose logarithm will be denoted by L, depends upon $p+q$ population parameters $\theta_1, \theta_2, \ldots, \theta_{p+q}$, assumed functionally independent. We shall assume that L and its partial derivatives with respect to the θ's satisfy some uniform continuity condition which allows differentiations with respect to the θ's to commute with integration over the sample space. We shall further assume that the second derivatives $\partial^2 L/\partial\theta_r^2$ are of order n, where n is related to the number of observations. These assumptions would usually be satisfied in practice. Let $\theta_r^{(0)}$ denote the true value of θ_r. We shall be concerned with testing the 'composite' hypothesis H_0 that $\theta_{p+1}, \theta_{p+2}, \ldots, \theta_{p+q}$ have specified values, which if H_0 is true we can take to be $\theta_{p+1}^{(0)}, \theta_{p+2}^{(0)}, \ldots, \theta_{p+q}^{(0)}$, while $\theta_1, \theta_2, \ldots, \theta_p$ are unspecified and unknown 'nuisance' parameters. The criterion obtained by taking minus twice the log likelihood ratio will be written as $2(L^{(p+q)} - L^{(p)})$, where $L^{(k)}$ denotes the result of maximizing L with respect to $\theta_1, \theta_2, \ldots, \theta_k$ and substituting true values for the remaining parameters.

We shall make full use of methods developed by Bartlett (1953*a*, *b*, 1955) in three papers on approximate confidence intervals. The expansion for $L^{(k)}$ which we now obtain follows very closely § 8 of the second of these papers (in future referred to as II), though our notation is different and we have taken in terms of higher order. We shall use the notation

$$l_r = \partial L/\partial\theta_r, \quad L_{rs} = \partial^2 L/\partial\theta_r\partial\theta_s, \quad L_{rsl} = \partial^3 L/\partial\theta_r\partial\theta_s\partial\theta_l, \quad \text{etc.},$$
$$\lambda_{rs} = E(L_{rs}), \quad \lambda_{rsl} = E(L_{rsl}), \quad \text{etc.},$$
$$l_{rs} = L_{rs} - \lambda_{rs}, \quad l_{rsl} = L_{rsl} - \lambda_{rsl}, \quad \text{etc.}$$

Then all λ's are, in general, of order n and the l's are random variates of order $\sqrt{n}$ with zero expectations.

Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ be the estimates inserted for $\theta_1, \theta_2, \ldots, \theta_k$ in L to produce $L^{(k)}$. The equations for determining the $\hat{\theta}_r$ may be written

$$0 = l_r + L_{rs}x_s + \tfrac{1}{2}L_{rst}x_s x_t + \tfrac{1}{6}L_{rstu}x_s x_t x_u + \ldots \quad (r = 1, 2, \ldots, k), \tag{1}$$

where $x_s = \hat{\theta}_s - \theta_s$, and where here and elsewhere, unless otherwise stated, the usual summation convention is employed. All suffices run from 1 to k. The inverse expansion of x_r in terms of the l_s is found, with a little algebra, to be

$$x_r = -L^{rs}l_s - \tfrac{1}{2}a_{rst}l_s l_t - \tfrac{1}{2}b_{rs,tu}l_s l_t l_u + \tfrac{1}{6}c_{rstu}l_s l_t l_u + \ldots, \tag{2}$$

where

$$a_{rst} = L^{rg}L^{sh}L^{ti}L_{ghi},$$
$$b_{rs,tu} = L^{rg}L^{sh}L^{ti}L^{uj}L^{vw}L_{ghv}L_{ijw},$$
$$c_{rstu} = L^{rg}L^{sh}L^{ti}L^{uj}L_{ghij},$$

and where $[L^{rs}]$ is the inverse matrix of $[L_{rs}]$.

We next write

$$L^{(k)} = L + l_r x_r + \tfrac{1}{2}L_{rs}x_r x_s + \tfrac{1}{6}L_{rst}x_r x_s x_t + \tfrac{1}{24}L_{rstu}x_r x_s x_t x_u + \ldots,$$

which in view of (1), is equivalent to

$$L - L^{(k)} = \tfrac{1}{2}L_{rs}x_r x_s + \tfrac{1}{3}L_{rst}x_r x_s x_t + \tfrac{1}{8}L_{rstu}x_r x_s x_t x_u + \ldots.$$

Substituting for the x_r in this, using (2), we have

$$2(L^{(k)} - L) = -L^{rs}l_r l_s - \tfrac{1}{3}a_{rst}l_r l_s l_t - \tfrac{1}{4}b_{rs,tu}l_r l_s l_t l_u + \tfrac{1}{12}c_{rstu}l_r l_s l_t l_u + \ldots.$$

Hence, including terms up to the fourth degree in the l's,

$$\begin{aligned} 2(L^{(k)} - L) = &-\lambda^{rs}l_r l_s - \tfrac{1}{3}\alpha_{rst}l_r l_s l_t + \lambda^{rt}\lambda^{su}l_r l_s l_{tu} - \tfrac{1}{4}\beta_{rs,tu}l_r l_s l_t l_u \\ &+ \tfrac{1}{12}\gamma_{rstu}l_r l_s l_t l_u + \lambda^{rv}\alpha_{stu}l_r l_s l_t l_{uv} - \tfrac{1}{3}\lambda^{ru}\lambda^{sv}\lambda^{tw}l_r l_s l_t l_{uvw} \\ &- \lambda^{ru}\lambda^{sv}\lambda^{tw}l_r l_s l_{tu} l_{vw}, \end{aligned} \tag{3}$$

where

$$\alpha_{rst} = \lambda^{rg}\lambda^{sh}\lambda^{ti}\lambda_{ghi},$$
$$\beta_{rs,tu} = \lambda^{rg}\lambda^{sh}\lambda^{ti}\lambda^{uj}\lambda^{vw}\lambda_{ghv}\lambda_{ijw},$$
$$\gamma_{rstu} = \lambda^{rg}\lambda^{sh}\lambda^{ti}\lambda^{uj}\lambda_{ghij},$$

and where $[\lambda^{rs}]$ is the inverse matrix of $[\lambda_{rs}]$. We have here taken the expansion as far as terms of order n^{-1}. As an explanation of the way in which (3) is obtained we note that

$$L^{rs} = \lambda^{rs} - \lambda^{rt}\lambda^{su}l_{tu} + \lambda^{ru}\lambda^{sv}\lambda^{tw}l_{tu}l_{vw} - \ldots,$$

and that

$$\begin{aligned} a_{rst}l_r l_s l_t &= \lambda^{ru}\lambda^{sv}\lambda^{tw}(\lambda_{uvw} + l_{uvw})l_r l_s l_t - 3\lambda^{rv}\lambda^{ug}\lambda^{sh}\lambda^{ti}\lambda_{ghi}l_r l_s l_t l_{uv} + \ldots \\ &= \alpha_{rst}l_r l_s l_t + \lambda^{ru}\lambda^{sv}\lambda^{tw}l_r l_s l_t l_{uvw} - 3\lambda^{rv}\alpha_{stu}l_r l_s l_t l_{uv} + \ldots. \end{aligned}$$

3. Various expectations

We now require the expectations of various products of the l's. These are readily obtained by the method given in § 2 of II. We shall suppose in future that H_0 is true and that each θ_r takes its true value $\theta_r^{(0)}$. Using the notation

$$(\lambda_{rs})_t = \partial\lambda_{rs}/\partial\theta_t, \qquad (\lambda_{rst})_u = \partial\lambda_{rst}/\partial\theta_u,$$
$$(\lambda_{rs})_{tu} = \partial^2\lambda_{rs}/\partial\theta_t\partial\theta_u, \quad \kappa_{(rs)(tu)} = E(l_{rs}l_{tu}),$$

we find that
$$E(l_r l_s) = -\lambda_{rs},$$
$$E(l_r l_{st}) = -\lambda_{rst} + (\lambda_{st})_r,$$
$$E(l_r l_{stu}) = -\lambda_{rstu} + (\lambda_{stu})_r,$$
$$E(l_r l_s l_t) = 2\lambda_{rst} - \sum_{(3)} (\lambda_{rs})_t,$$
$$E(l_r l_s l_{tu}) = \lambda_{rstu} - (\lambda_{stu})_r - (\lambda_{rtu})_s + (\lambda_{tu})_{rs} - \kappa_{(rs)(tu)},$$
$$E(l_r l_s l_t l_u) = \sum_{(3)} (\lambda_{rs}\lambda_{tu}) - 3\lambda_{rstu} + 2\sum_{(4)} (\lambda_{rst})_u - \sum_{(6)} (\lambda_{rs})_{tu} + \sum_{(3)} \kappa_{(rs)(tu)},$$

where, for example,
$$\sum_{(3)} (\lambda_{rs})_t \equiv (\lambda_{rs})_t + (\lambda_{rt})_s + (\lambda_{st})_r.$$

Other expectations, in which we neglect terms of order n, are given by

$$E(l_r l_s l_t l_{uv}) = \sum_{(3)} \{\lambda_{ruv} - (\lambda_{uv})_r\} \lambda_{st} \quad (r, s, t \text{ permuted}),$$
$$E(l_r l_s l_t l_{uvw}) = \sum_{(3)} \{\lambda_{ruvw} - (\lambda_{uvw})_r\} \lambda_{st} \quad (r, s, t \text{ permuted}),$$
$$E(l_r l_s l_{tu} l_{vw}) = \{\lambda_{rtu} - (\lambda_{tu})_r\}\{\lambda_{svw} - (\lambda_{vw})_s\} + \{\lambda_{rvw} - (\lambda_{vw})_r\}\{\lambda_{stu} - (\lambda_{tu})_s\} - \lambda_{rs}\kappa_{(tu)(vw)}.$$

By making use of these results we find the expectation of $2(L^{(k)} - L)$, as given by (3), to be $k + \epsilon_k + O(n^{-2})$, where ϵ_k is of order n^{-1} and is given by

$$\epsilon_k = \lambda^{rs}\lambda^{tu}\{\tfrac{1}{4}\lambda_{rstu} - (\lambda_{rst})_u + (\lambda_{rt})_{su}\} - \lambda^{rs}\lambda^{tu}\lambda^{vw}\{\tfrac{1}{6}\lambda^2_{rtv}\lambda_{suw} + \tfrac{1}{4}\lambda_{rtu}\lambda_{svw} - \lambda_{rtv}(\lambda_{suw})_u - \lambda_{rtu}(\lambda_{sw})_v + (\lambda_{rt})_v(\lambda_{sw})_u + (\lambda_{rt})_u(\lambda_{sw})_v\}. \quad (4)$$

Here, as before, all suffices are summed over the values 1 to k. The expectation of the criterion $2(L^{(p+q)} - L^{(p)})$ may be written as
$$q + \epsilon_{p+q} - \epsilon_p + O(n^{-2}).$$

In order to simplify the ensuing algebra we shall now assume, which can be done without loss of generality, that the parameters were originally chosen such that, for all unequal r and s, $\lambda_{rs} = 0 = \lambda^{rs}$ (when the θ's are given their true values). We can then write, discarding temporarily the summation convention,

$$\epsilon_k = \sum_{r,s=1}^{k} \{\tfrac{1}{4}\lambda_{rrss} - (\lambda_{rrs})_s + (\lambda_{rs})_{rs}\}/(\lambda_{rr}\lambda_{ss}) + \sum_{r,s,t=1}^{k} \{\tfrac{1}{6}\lambda^2_{rst} + \tfrac{1}{4}\lambda_{rss}\lambda_{rtt} - \lambda_{rst}(\lambda_{rs})_t - \lambda_{rss}(\lambda_{rt})_t + (\lambda_{rs})_t(\lambda_{rt})_s + (\lambda_{rs})_s(\lambda_{rt})_t\}/(-\lambda_{rr}\lambda_{ss}\lambda_{tt}). \quad (5)$$

The further simplification will be made, again without loss of generality, of putting, for all r, $\lambda_{rr} = -n$, $\lambda^{rr} = -n^{-1}$ (for true values of the θ's). This enables us to write, in place of (3),

$$2(L^{(k)} - L) = \frac{1}{n} l_r l_r + \frac{1}{3n^3}\lambda_{rst} l_r l_s l_t + \frac{1}{n^2} l_r l_s l_{rs} + \frac{1}{12n^4}\lambda_{rstu} l_r l_s l_t l_u + \frac{1}{4n^5}\lambda_{rsv}\lambda_{tuv} l_r l_s l_t l_u + \frac{1}{n^4}\lambda_{rsu} l_r l_s l_t l_{tu} + \frac{1}{3n^3} l_r l_s l_t l_{rst} + \frac{1}{n^3} l_r l_s l_{rt} l_{st}, \quad (6)$$

ignoring terms of more than the fourth degree in the l's. A few words of explanation are desirable at this point. While the notation λ_{rs} is unambiguous, the meaning of λ^{rs} depends, in general, on the order k of the matrix which is being inverted. Nevertheless, if at the outset we choose the θ's such that $\lambda_{rr} = -n$ and $\lambda_{rs} = 0$ $(r \neq s)$ for $r, s = 1, 2, \ldots, p+q$, then $\lambda^{rr} = -n^{-1}$ and $\lambda^{rs} = 0$ whatever the value of k. We can, for example, make the choice of

θ's by means of a certain linear transformation. If the original set of θ's do not satisfy the above conditions we may replace each θ_r by $\theta'_r = \sum_{s \geqslant r} a_{rs}\theta_s$, with suitable values of the a_{rs}. By this means we obtain a set of θ's which makes (6) true for all k. These remarks are relevant to the definition of the quantities m_r^2 in the next section.

4. Joint cumulants of the m_r

Now define quantities $m_1^2, m_2^2, \ldots, m_{p+q}^2$ by

$$m_1^2 = 2(L^{(1)} - L),$$
$$m_r^2 = 2(L^{(r)} - L^{(r-1)}) \quad (r > 1).$$

Then
$$2(L^{(k)} - L) = \sum_{r=1}^{k} m_r^2,$$

and the criterion in which we are interested may be written as $\sum_{r=p+1}^{p+q} m_r^2$. It is readily found that m_r is given (with suitable choice of sign) as far as terms of the second degree in the l's by

$$m_r\sqrt{n} = l_r + \frac{1}{6n^2}\lambda_{rrr}l_r^2 + \frac{1}{2n^2}l_r\sum_{s<r}\lambda_{rrs}l_s + \frac{1}{2n^2}\sum_{s,t<r}\lambda_{rst}l_sl_t + \frac{1}{2n}l_rl_{rr} + \frac{1}{n}\sum_{s<r}l_sl_{rs}, \tag{7}$$

where we drop the summation convention, and continue to do so until further notice.

We now establish certain results concerning the joint moments and cumulants of the m_r. We shall denote these by $\mu_r, \mu_{rs}, \mu_{rst}, \ldots$ (with or without dashes) and by $\kappa_r, \kappa_{rs}, \kappa_{rst}, \ldots$. In our notation $\mu_{rs} = \kappa_{rs}$ is, for example, the covariance between m_r and m_s. For the mean of m_r we have

$$\kappa_r\sqrt{n} = \mu'_r\sqrt{n} = \frac{1}{6n}\lambda_{rrr} + \frac{1}{2n}\sum_{s<r}\lambda_{rss} + \frac{1}{2n}\{-\lambda_{rrr} + (\lambda_{rr})_r\} + \frac{1}{n}\sum_{s<r}\{-\lambda_{rss} + (\lambda_{rs})_s\} + O(n^{-1}).$$

Hence
$$\mu'_r = n^{-\frac{1}{2}}\{-\tfrac{1}{3}\lambda_{rrr} + \tfrac{1}{2}(\lambda_{rr})_r\} + n^{-\frac{3}{2}}\sum_{s<r}\{-\tfrac{1}{2}\lambda_{rss} + (\lambda_{rs})_s\} + O(n^{-\frac{5}{2}}).$$

Since the expectation of $\sum_{r=1}^{k} m_r^2$ is $k + \epsilon_k + O(n^{-2})$, where ϵ_k is given by (5) with $\lambda_{rr} = -n$, we have $\mu'_{rr} = 1 + \epsilon_r - \epsilon_{r-1} + O(n^{-2})$. It is unnecessary for us to evaluate μ'_{rs} $(r \neq s)$. We need only observe that both μ'_{rs} and $\mu_{rs} = \kappa_{rs}$ are of order n^{-1}.

For the third-order moments and cumulants we have, first, that $n^{\frac{3}{2}}\mu'_{rrr}$ is the expectation of

$$l_r^3 + \frac{1}{2n^2}\lambda_{rrr}l_r^4 + \frac{3}{2n^2}l_r^3\sum_{s<r}\lambda_{rrs}l_s + \frac{3}{2n^2}l_r^2\sum_{s,t<r}\lambda_{rst}l_sl_t + \frac{3}{2n}l_r^3l_{rr} + \frac{3}{n}l_r^2\sum_{s<r}l_sl_{rs} + \ldots,$$

and thus
$$\mu'_{rrr} = n^{-\frac{3}{2}}\{-\lambda_{rrr} + \tfrac{3}{2}(\lambda_{rr})_r\} + 3n^{-\frac{3}{2}}\sum_{s<r}\{-\tfrac{1}{2}\lambda_{rss} + (\lambda_{rs})_s\} + O(n^{-\frac{5}{2}})$$
$$= 3\mu'_r + O(n^{-\frac{1}{2}}).$$

Similarly we find
$$\mu'_{rrs} = \mu'_s + O(n^{-\frac{3}{2}}) \quad (r \neq s),$$
$$\mu'_{rst} = O(n^{-\frac{3}{2}}) \quad (r, s, t \text{ all unequal}).$$

Hence, for all values of r, s and t, $\kappa_{rst} = \mu_{rst} = O(n^{-\frac{3}{2}})$.

We must next show that the fourth-order cumulants κ_{rrrr} and κ_{rrss} are of order n^{-2}. The method employed for doing this involves exceedingly complicated and laborious algebra, and has the disadvantage of making the final result appear miraculous! We have not,

however, been able to discover a better method. We shall find the covariance of $\sum_{i=1}^{k} m_i^2$, as given by (6), with $\sum_{s=1}^{r} m_s^2 (r \geqslant k)$, a similar expression in which, however, all suffices are summed over the values 1 to r. We must retain all terms of order n^{-1}. It is hardly desirable to reproduce all the algebra, but it may help to give a few illustrations. We reintroduce the summation convention with the understanding that the suffices g, h, i and j are to be summed from 1 to k, while the suffices s, t and u are summed from 1 to r ($\geqslant k$). Four of the terms which arise are:

$$\frac{1}{n^2}\operatorname{cov}\{(l_s l_s), (l_i l_i)\}$$

$$= 2k + \frac{1}{n^2}\{-3\lambda_{ssii} + 4(\lambda_{ssi})_i + 4(\lambda_{sii})_s - (\lambda_{ss})_{ii} - (\lambda_{ii})_{ss} - 4(\lambda_{si})_{si} + \kappa_{(ss)(ii)} + 2\kappa_{(si)(si)}\},$$

$$\frac{1}{3n^4}\lambda_{ghi}\operatorname{cov}\{(l_s l_s), (l_g l_h l_i)\}$$

$$= \frac{1}{n^3}\lambda_{ghi}\{4\lambda_{ghi} - 6(\lambda_{gh})_i\} + \frac{1}{n^3}\lambda_{iij}\{2\lambda_{ssj} - 2(\lambda_{sj})_s - (\lambda_{ss})_j\},$$

$$\frac{1}{n^4}\lambda_{ghj}\operatorname{cov}\{(l_s l_s), (l_g l_h l_i l_{ij})\}$$

$$= \frac{8}{n^3}\lambda_{ghi}\{-\lambda_{ghi} + (\lambda_{gh})_i\} + \frac{4}{n^3}\lambda_{ggi}\{-\lambda_{hhi} + (\lambda_{hi})_h\},$$

$$\frac{1}{3n^5}\lambda_{stu}\operatorname{cov}\{(l_s l_t l_u), (l_i l_j l_{ij})\}$$

$$= \frac{1}{n^3}\lambda_{ssi}\{-\lambda_{tti} + (\lambda_{it})_t\} + \frac{2}{n^3}\lambda_{ssj}\{-\lambda_{iij} + (\lambda_{ij})_i\} + \frac{2}{n^3}\lambda_{sij}\{-\lambda_{sij} + (\lambda_{ij})_{sj}\}.$$

There are altogether nineteen such terms. Their sum is found to reduce to

$$2k + \frac{1}{n^2}\{\lambda_{iijj} - 4(\lambda_{iij})_j + 4(\lambda_{ij})_{ij}\} + \frac{1}{n^3}\{\tfrac{2}{3}\lambda_{ggi}\lambda_{ghi} + \lambda_{ggi}\lambda_{hhi}$$

$$- 4\lambda_{ghi}(\lambda_{gi})_h - 4\lambda_{ggi}(\lambda_{hi})_h + 4(\lambda_{gi})_h(\lambda_{hi})_g + 4(\lambda_{gi})_g(\lambda_{hi})_h\}.$$

Hence the covariance between $\sum_{s=1}^{r} m_s^2$ and $\sum_{i=1}^{k} m_i^2$ $(r \geqslant k)$ is $2k + 4\epsilon_k + O(n^{-2})$. From this it is easily deduced that

$$\operatorname{var}(m_r^2) = 2 + 4(\epsilon_r - \epsilon_{r-1}) + O(n^{-2}),$$

$$\operatorname{cov}(m_r^2, m_s^2) = O(n^{-2}) \quad (r \neq s),$$

and that

$$\mu'_{rrrr} = 3\mu'^2_{rr} + O(n^{-2}), \qquad \kappa_{rrrr} = O(n^{-2}),$$

$$\mu'_{rrss} = \mu'_{rr}\mu'_{ss} + O(n^{-2}), \quad \kappa_{rrss} = O(n^{-2}) \quad (r \neq s).$$

Since the leading term of m_r is $l_r/\sqrt{n}$, it is easy to show that all other fourth-order cumulants are certainly $O(n^{-1})$ or less. Similarly, regarding the cumulants of higher order we need only remark that those of the fifth order are not more than $O(n^{-\frac{3}{2}})$, those of the sixth order not more than $O(n^{-2})$, and so on.

5. Approximate distribution of the criterion

Now consider the expectation of $\prod_{i=p+1}^{p+q} (m_i^{2\rho_i})$, where the ρ_i are non-negative integers, and suppose that it is expressed in terms of the joint cumulants of the m_i. Then, bearing in mind the orders of magnitude of these cumulants, it is clear that the only terms which matter are, first, that in $\prod_i (\kappa_{ii}^{\rho_i})$, which has a coefficient of $\prod_i \left\{\frac{(2\rho_i)!}{2^{\rho_i}\rho_i!}\right\}$, and, secondly, those in which *one* of the powers $\kappa_{ii}^{\rho_i}$ is replaced by $\kappa_{ii}^{\rho_i-1}\kappa_i^2$, with a coefficient ρ_i times as large. All other terms are $O(n^{-2})$ or less. Since $\mu'_{ii} = \kappa_{ii} + \kappa_i^2$, it follows that, neglecting quantities of order n^{-2}, the expectation of $\prod_i (m_i^{2\rho_i})$ is given by

$$\prod_i \left\{\frac{(2\rho_i)!}{2^{\rho_i}\rho_i!}\mu'^{\rho_i}_{ii}\right\}.$$

In view of this it is clear that, to the same order of approximation, the moments of the criterion $\sum_{i=p+1}^{p+q} m_i^2$ are the same as those of

$$a_{p+1}z_1 + a_{p+2}z_2 + \ldots + a_{p+q}z_q,$$

where $a_i = \mu'_{ii} = 1 + \epsilon_i - \epsilon_{i-1}$, and where $z_1, z_2, \ldots, z_q$ are independent χ^2 variates each having one degree of freedom. So the ρth cumulant of the criterion (assuming n large compared with ρ) is

$$\begin{aligned} 2^{\rho-1}(\rho-1)! \sum_{i=p+1}^{p+q} a_i^\rho &= 2^{\rho-1}(\rho-1)!\,\{q + \rho(\epsilon_{p+q} - \epsilon_p)\} + O(n^{-2}) \\ &= 2^{\rho-1}(\rho-1)!\,q\left\{1 + \frac{1}{q}(\epsilon_{p+q} - \epsilon_p)\right\}^\rho + O(n^{-2}). \end{aligned}$$

Hence, finally, either $\quad 2q(L^{(p+q)} - L^{(p)})/(q + \epsilon_{p+q} - \epsilon_p)$

or $\quad 2\left\{1 - \frac{1}{q}(\epsilon_{p+q} - \epsilon_p)\right\}(L^{(p+q)} - L^{(p)})$

has the same moments as χ^2 with q degrees of freedom, neglecting quantities of order n^{-2}.

Though the quantity ϵ_k is given either by (4) or by (5), these expressions would not as a rule be of much practical use, particularly if the number of unknown parameters were large. In particular cases it would generally be much easier to find $q + \epsilon_{p+q} - \epsilon_p$ by evaluating directly the expectation of the criterion as far as terms of order n^{-1}. This method was adopted in a previous paper (Lawley, 1956) when testing hypotheses regarding the latent roots of a covariance matrix. It can be shown that both the hypotheses and the criteria employed to test them were of the same type as those considered in this paper. In general $\epsilon_{p+q} - \epsilon_p$ will be a function of the unknown parameters, but the substitution of estimated values for these will clearly not affect the order of the approximation.

We end with three examples. In both of the first two the criterion belongs to the class considered by Box, and the results are not new, but our main object here is to provide some slight verification of expression (5).

Example I. Suppose that we have a random sample of size n from a normal population. We consider the well-known problem of testing the hypothesis that the mean θ_2 has a specified value, while the variance θ_1 is unknown. In this case

$$L = -\tfrac{1}{2}n\log\theta_1 - \frac{1}{2\theta_1}\{n(\bar{x} - \theta_2)^2 + (n-1)s^2\},$$

where $\bar{x}$ is the sample mean and s^2 is the usual estimate of θ_1. The criterion is

$$2(L^{(2)}-L^{(1)}) = n\log\left(1+\frac{t^2}{n-1}\right),$$

where $t = (\bar{x}-\theta_2)\sqrt{n}/s$.

Differentiating L we find, omitting various zero quantities,

$$\lambda_{11} = -\frac{n}{2\theta_1^2}, \quad \lambda_{22} = -\frac{n}{\theta_1}, \quad \lambda_{12} = 0,$$

$$\lambda_{122} = \frac{n}{\theta_1^2} = (\lambda_{22})_1, \quad \lambda_{1122} = -\frac{2n}{\theta_1^3} = (\lambda_{122})_1.$$

Since $\lambda_{12} = 0$, we are able to use (5) and hence obtain $1+\epsilon_2-\epsilon_1 = 1+3/(2n)$, which is the right value for the expectation of the criterion. The corrected criterion is therefore

$$\chi^2_{(1)} = (n-\tfrac{3}{2})\log\left(1+\frac{t^2}{n-1}\right),$$

which may be expanded as $$t^2\left(1-\frac{1+t^2}{2\nu}+\frac{3t^2+4t^4}{12\nu^2}-\ldots\right),$$

where $\nu = n-1$. This happens to give the 5 and 1% significance levels fairly accurately even for $n = 6$, and it agrees, as far as the term in ν^{-1}, with the correct expansion, which is known to be

$$t^2\left(1-\frac{1+t^2}{2\nu}+\frac{3+7t^2+8t^4}{24\nu^2}-\ldots\right).$$

This may be obtained by inverting the expansion given by Fisher (1941).

Example II. As a second example, suppose that we have k independent variance estimates s_i^2 with expectations σ_i^2 and degrees of freedom n_i $(i = 1, 2, \ldots, k)$. Then (assuming normal parent populations) we may take L to be

$$-\frac{1}{2}\sum_{i=1}^{k} n_i\{\log(\sigma_i^2)+s_i^2/\sigma_i^2\}.$$

We assume that the n_i are all of the same order of magnitude and that quantities of order n_i^{-2} can be ignored. The hypothesis H_0 to be tested is that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, and our criterion is $2(L^{(k)}-L^{(1)})$, where $L^{(k)}$ and $L^{(1)}$ have to be determined.

For $L^{(1)}$ it is convenient to define the parameters θ_i by

$$\theta_1 = \sigma_1^2, \quad \theta_i = \sigma_i^2-\sigma_1^2 \quad (i>1).$$

Then, assuming H_0 to be true, the true values of $\theta_2, \ldots, \theta_k$ are all zero, and L becomes

$$-\tfrac{1}{2}n(\log\theta_1+s^2/\theta_1),$$

where $$n = \sum_{i=1}^{k} n_i, \quad s^2 = \sum_i (n_i s_i^2)/n.$$

Maximizing this with respect to θ_1, we obtain

$$L^{(1)} = -\tfrac{1}{2}n\{\log(s^2)+1\}, \quad \hat{\theta}_1 = s^2.$$

Differentiation of L with respect to θ_1 gives

$$\lambda_{11} = -\frac{n}{2\theta_1^2}, \quad \lambda_{111} = \frac{2n}{\theta_1^3}, \quad (\lambda_{11})_1 = \frac{n}{\theta_1^3},$$

$$\lambda_{1111} = -\frac{9n}{\theta_1^4}, \quad (\lambda_{111})_1 = -\frac{6n}{\theta_1^4}, \quad (\lambda_{11})_{11} = -\frac{3n}{\theta_1^4}.$$

Hence use of (5) gives the expectation of $2(L^{(1)}-L)$ as $1+\epsilon_1 = 1+1/(3n)$.

Since $L^{(k)}$ is the result of maximizing L with respect to all the parameters it is in this case more convenient to define θ_i as σ_i^2, for all i. We then have

$$L^{(k)} = -\tfrac{1}{2}\sum_i n_i \log(s_i^2) - \tfrac{1}{2}n,$$

and also $$\lambda_{ii} = -\frac{n_i}{2\theta_i^2}, \quad \lambda_{iii} = \frac{2n_i}{\theta_i^3}, \quad \text{etc.}$$

Hence, from (5), the expectation of $2(L^{(k)}-L)$ is

$$k+\epsilon_k = k+\tfrac{1}{3}\sum_i\left(\frac{1}{n_i}\right).$$

Thus the criterion, which is $$n\log(s^2) - \sum_i n_i \log(s_i^2),$$

has an expectation of $(k-1)+\frac{1}{3}\left(\sum_i \frac{1}{n_i} - \frac{1}{n}\right)$. The χ^2 test is improved by using

$$1+\frac{1}{3(k-1)}\left(\sum_i \frac{1}{n_i} - \frac{1}{n}\right)$$

as a divisor for the criterion, as first established by Bartlett (1937).

Example III. Lastly we consider the hypothesis that the correlation coefficient ρ $(=\theta_3)$ in a bivariate normal distribution has a specified value, the two standard deviations θ_1 and θ_2 being nuisance parameters. This example was considered by Bartlett in the third of his papers, already referred to, on approximate confidence intervals. We shall take the approximation a further stage. If r, s_1 and s_2 are the usual estimates of ρ, θ_1 and θ_2 respectively, obtained from a sample of size $n+1$, then, omitting a constant, we have

$$L = -n\log\{\theta_1\theta_2\sqrt{(1-\rho^2)}\} - \frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\theta_1^2} - \frac{2\rho r s_1 s_2}{\theta_1\theta_2} + \frac{s_2^2}{\theta_2^2}\right\}.$$

Our criterion is easily found to be given by

$$2(L^{(3)}-L^{(2)}) = n\log\left\{\frac{(1-r\rho)^2}{(1-\rho^2)(1-r^2)}\right\}.$$

In this case it would be exceedingly laborious to calculate the expectation by differentiating L, and we therefore find it more directly. Expansion of the above expression in powers of $x = (r-\rho)/(1-\rho^2)$ gives

$$n\log\{1+x^2+2\rho x^3+(1+3\rho^2)x^4+\ldots\} = n\{x^2+2\rho x^3+\tfrac{1}{2}(1+6\rho^2)x^4+\ldots\}.$$

Making use of results obtained by Hotelling (1953) we have, neglecting quantities of order n^{-3},

$$E(x^2) = \frac{1}{n}+\frac{23\rho^2}{4n^2}, \quad E(x^3) = -\frac{15\rho}{2n^2}, \quad E(x^4) = \frac{3}{n^2}.$$

Hence the required expectation is $1+(6-\rho^2)/(4n)+O(n^{-2})$. For an improved criterion we use, in place of n, the multiplying factor $n-\frac{1}{4}(6-\rho^2)$.

We may expand in powers of $z-\zeta$, instead of x, where

$$z=\tfrac{1}{2}\log\frac{1+r}{1-r},\quad \zeta=\tfrac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

To the usual order of approximation we then have that

$$\{n-\tfrac{1}{4}(6-\rho^2)\}\{(z-\zeta)^2-\tfrac{1}{6}(z-\zeta)^4\}$$

is distributed as χ^2 with one degree of freedom; and it may easily be verified that the variate $(z-\zeta)-\frac{1}{12}(z-\zeta)^3$ has cumulants given by

$$\kappa_1=\frac{\rho}{2n}+O(n^{-2}),\quad \kappa_2=\frac{1}{n}+\frac{1}{2n^2}(3-\rho^2)+O(n^{-3}),$$
$$\kappa_3=O(n^{-3}),\qquad \kappa_4=O(n^{-4}).$$

This could be used to provide confidence limits for ρ, but no practical improvement on the ordinary use of z for this purpose would be achieved.

My thanks are due to Mr D. V. Lindley for his helpful suggestions for clarifying the argument in various places.

REFERENCES

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc.* A, **160**, 268.

BARTLETT, M. S. (1953*a*). Approximate confidence intervals. I. *Biometrika*, **40**, 12.

BARTLETT, M. S. (1953*b*). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, **40**, 306.

BARTLETT, M. S. (1954). A note on the multiplying factors in various χ^2 approximations. *J. R. Statist. Soc.* B, **16**, 296.

BARTLETT, M. S. (1955). Approximate confidence intervals. III. A bias correction. *Biometrika*, **42**, 201.

BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the d test of significance. *Ann. Eugen., Lond.*, **11**, 141.

HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *J. R. Statist. Soc.* B, **15**, 193.

LAWLEY, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128.

NEYMAN, J. & PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20A**, 175 and 263.

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60.

ON THE ACCURACY OF WEIGHTED MEANS AND RATIOS

By G. S. JAMES

University of Leeds

1. Introduction

The basic problem we consider is as follows. Suppose $x_1, \ldots, x_k$ are k quantities, derived from observations, which are independently and normally distributed about the same mean μ, but with possibly different variances $\lambda_1\sigma_1^2, \ldots, \lambda_k\sigma_k^2$. The positive constants λ_i are known. The σ_i^2 are unknown, but estimates s_i^2 are available which are distributed independently of each other and of the x_i in the usual mean-square forms with ν_i degrees of freedom (ν_i known). We require to find confidence limits for μ. For example, the x_i may be the means of k samples (sizes $n_1, \ldots, n_k$) drawn from normal populations having the same mean μ, but perhaps different variances $\sigma_1^2, \ldots, \sigma_k^2$. The s_i^2 are then the sample variances defined with degrees of freedom $\nu_i = n_i - 1$, and the λ_i are the reciprocals of the n_i. Again, the x_i might be the estimated slopes of k regression lines, the true lines being known to have the same slope μ, but the residual variances σ_i^2 being possibly unequal. The s_i^2 are now the usual estimates derived from the deviations from the fitted lines, while the λ_i are the reciprocals of the sums of squares of deviations of the independent variate.

Let ω_i denote $1/(\lambda_i\sigma_i^2)$, the reciprocal-variance or *weight* of x_i, and let w_i denote $1/(\lambda_i s_i^2)$. Write ω for $\Sigma\omega_i$ and w for Σw_i. Now if the true weights ω_i (or their ratios) were known, then the quantity $\bar{x} = \Sigma\omega_i x_i/\omega$ would provide an estimate of μ of maximal accuracy (its variance being ω^{-1}). It therefore seems reasonable, in the absence of any firm knowledge of the true weights, to use the estimate $\hat{x} = \Sigma w_i x_i/w$ in its place. Now although $\omega^{\frac{1}{2}}(\bar{x} - \mu)$ has a standard normal distribution, that of $u = w^{\frac{1}{2}}(\hat{x} - \mu)$ has a form which depends on the ratios of the ω_i, which are unknown. Thus tables of this distribution, even if available, would be of no assistance in finding confidence limits for μ. Nevertheless, it may be possible to find a function $u(r_1, \ldots, r_k)$, or $u(r)$ for short, of the ratios $r_i = w_i/w$ ($\Sigma r_i = 1$), which is such that

$$\Pr[|u| \leqslant u(r)] = P, \tag{1·1}$$

either exactly or approximately, where P is the required confidence coefficient. Confidence limits for μ are then $\hat{x} \pm u(r)/w^{\frac{1}{2}}$. Of course $u(r)$ will depend in addition on P and on the degrees of freedom ν_i. It is not known for certain whether the functional equation (1·1) does or does not possess an *exact* solution, but in this paper we present a solution which is asymptotically correct when all the degrees of freedom are large. More specifically we give a function $u(r)$ for which

$$\Pr[|u| \leqslant u(r)] = P + O(\nu^{-4}) \tag{1·2}$$

and tabulate it for the case $k = 2$.

The ordinary large-sample approximation would be to take $u(r)$ equal to the appropriate point of the standard normal distribution. This is also asymptotically correct, but satisfies (1·2) with ν^{-1} in place of ν^{-4}.

We also show how the tables of this paper, and certain others originating from work by B. L. Welch, may be used to provide confidence limits for certain quantities which are estimated as ratios.

2. General discussion: Yates's solution

In a sense the present problem is dual to that discussed by Welch (1947*a*, *b*) and by Trickett & Welch (1954) (see also further references below). Their problem was that of finding confidence limits for a previously specified linear combination, $\eta = \Sigma\beta_i\mu_i$, of k quantities μ_i (not here assumed equal), the assumptions being otherwise the same as before. (In particular, confidence limits were desired for the difference between the means of two normal populations, whose variances were not assumed to be equal.) More precisely, if $y = \Sigma\beta_i x_i$, $a = \Sigma a_i = \Sigma\beta_i^2\lambda_i s_i^2$ (the estimated variance of y), $c_i = a_i/a$ ($\Sigma c_i = 1$) and $v = a^{-\frac{1}{2}}(y-\eta)$, then the problem was to find a function $v(c_1, \ldots, c_k) = v(c)$ which was such that

$$\Pr[|v| \leqslant v(c)] = P. \tag{2·1}$$

Functions $v(c)$ have been given which satisfy this equation within terms of order ν^{-3} (Welch, 1947*a*, *b*), and within terms of order ν^{-5} (Aspin, 1948). Tables (for the case $k=2$) have been given by Aspin (1949) and by Trickett, Welch & James (1956) (see also Pearson & Hartley, 1954).

Now a different solution to the problem of finding limits for the difference of two means was proposed by Behrens (1929). This has been rederived by Fisher (1935, 1939, 1941) and Yates (1939) using the theory of fiducial distributions of population parameters, and by Jeffreys (1940) using an inverse probability approach. Tables (again for the case $k=2$) have been given by Sukhatme (1938) and by Fisher (1941) (see also Fisher & Yates, 1953). The argument using fiducial distributions is as follows. Let

$$t_i = (x_i - \mu_i)/(\lambda_i^{\frac{1}{2}} s_i) = \beta_i(x_i - \mu_i)/a_i^{\frac{1}{2}}. \tag{2·2}$$

Then the t_i have independent Student distributions with ν_i degrees of freedom. Equation (2·2) is equivalent to

$$\mu_i = x_i - a_i^{\frac{1}{2}} t_i/\beta_i, \tag{2·3}$$

and so

$$\eta = y - \Sigma a_i^{\frac{1}{2}} t_i. \tag{2·4}$$

It is now asserted that given a *unique* set of values x_i and s_i^2, (2·3) gives the distribution (in the fiducial sense) of the parameters μ_i. That is to say, if t_{iP} denotes the value of $|t|$, based on ν_i degrees of freedom, which is only exceeded with probability $(1-P)$, then the *fiducial* probability that μ_i lies between the (fixed) limits $x_i \pm a_i^{\frac{1}{2}} t_{iP}/\beta_i$ is numerically equal to P. That is to say the fiducial distribution of μ_i is obtained from that of t_i, using (2·3) and treating x_i and a_i as constants: it is merely a transformed t-distribution. Likewise the fiducial distribution of η, given the unique samples, is said to be that of the linear combination of t-variates given by (2·4), with y and the a_i treated as constants. Now

$$v = a^{-\frac{1}{2}}(y-\eta) = \Sigma c_i^{\frac{1}{2}} t_i. \tag{2·5}$$

Thus if $d = d(c)$ denotes that number which is such that (for constant c_i)

$$\Pr[|\Sigma c_i^{\frac{1}{2}} t_i| \leqslant d(c) \,|\, c] = P, \tag{2·6}$$

then it is asserted that

$$\mathrm{Fp}[|v| \equiv a^{-\frac{1}{2}}|\eta - y| \leqslant d(c)] = P, \tag{2·7}$$

where 'Fp' denotes the fiducial probability† of the relation following. This is a fiducial statement about η given a unique set of values x_i and s_i^2. It is *not* asserted, and it is not true in general, that

$$\Pr[|v| \equiv a^{-1}|y-\eta| \leqslant d(c)] = P, \tag{2·8}$$

where this statement is interpreted in the ordinary way (η a constant; y, a and the c_i having their ordinary direct probability distributions). It is held by those who use the fiducial argument that the fact that (2·8) is untrue is completely irrelevant, and that when, as in many other procedures based on normal distributions, statements equivalent to (2·7) and (2·8) are both true (each with its own interpretation), this is a confusing accident.

However, this does not seem to be the place to argue about the two points of view. My purpose in mentioning the fiducial argument is to recall that Yates (1939) showed that it leads to a solution of the weighted mean problem which can be applied using the same tables as are needed for the Behrens–Fisher problem.‡ For if the r_i can be regarded as constants, then

$$u = w^{\frac{1}{2}}(\hat{x}-\mu) = \Sigma r_i^{\frac{1}{2}} t_i \tag{2·9}$$

has the same distribution as v of (2·5), except that the quantities c_i are replaced by the r_i. Thus

$$\mathrm{Fp}[|u| \equiv w^{\frac{1}{2}}|\mu-\hat{x}| \leqslant d(r)] = P. \tag{2·10}$$

This equation gives limits, in Fisher's fiducial sense, for the parameter μ. The duality between the original Behrens–Fisher problem and the weighted mean problem with $k=2$ is particularly striking, because with β_1 and $\beta_2 = \pm 1$ we have $r_1 = c_2$ and $r_2 = c_1$. In the Sukhatme–Fisher tables, where the quantity $\theta = \sin^{-1} c_1^{\frac{1}{2}}$ is used as an argument instead of $c = c_1$, the change merely amounts to using $(\frac{1}{2}\pi - \theta)$ in place of θ.

If confidence limits in the ordinary sense of equation (1·1) are required, the tables of Welch's $v(c)$ cannot be adapted in a similar manner, and completely new ones, given later, are necessary. Even with the two sets of tables available the duality (if indeed that be the correct word) is very one-sided; for the u-problem is fundamentally more complicated than Welch's v-problem. This is because in the latter the combination occurring in the numerator is $\beta_1 x_1 + \beta_2 x_2$, the β_i being constants, whereas in the former it is $w_1 x_1 + w_2 x_2$, the w_i being random variables. In fiducial theory the w_i are also (at a certain stage) treated as constants, thus leading to an almost complete duality between the two problems.

3. The asymptotic solution

We suppose that we have quantities $x_1, \ldots, x_k$ which are normally distributed with the same mean μ but with possibly different variances $\alpha_1 = 1/\omega_1, \ldots, \alpha_k = 1/\omega_k$. Also that $a_1 = 1/w_1, \ldots, a_k = 1/w_k$ are estimates of the α_i, $\nu_i a_i/\alpha_i$ being distributed as χ_i^2 with ν_i degrees of freedom, the ν_i being known positive integers. The x_i and a_i distributions are supposed

† I am assuming that an expression such as the left-hand side of (2·7) is capable of a unique interpretation. Fisher has laid great emphasis on the theoretical necessity of defining the fiducial distributions of parameters in a unique fashion, but even when one only makes use of sample statistics jointly sufficient for the parameters, it sometimes seems to be possible to produce more than one fiducial distribution for the same quantity. The interested reader may consult Creasy (1954) (including the Discussion held at the end of the Symposium), and Mauldon (1955).

‡ As in the solution discussed later, relatively small amounts of information on $\sigma_1^2, \ldots, \sigma_k^2$ contained in the contrasts between $x_1, \ldots, x_k$ are neglected, this being not theoretically in accord with the conditions laid down by Fisher for the derivation of fiducial distributions; but in practice this can be neglected for the time being.

all to be mutually independent. No prior knowledge of μ or the α_i is assumed, and it is required to find confidence limits for μ with confidence coefficient (asymptotically) equal to P. Let $\hat{x}=\Sigma w_i x_i/w$, where w denotes Σw_i. Then the problem will be solved if we can find a function $u(a_1, \ldots, a_k)=u(a)$ of the sample variances (and of P and the ν_i) which is such that (asymptotically)

$$\Pr[w^{\frac{1}{2}} \mid \hat{x}-\mu \mid \leqslant u(a)]=P. \tag{3·1}$$

As remarked in § 1, $u(a)$ actually depends only on the *ratios* of the sample variances or weights, and can be written $u(r)$, where $r_i=w_i/w$, but the above form is more convenient for present purposes. It must be remembered that both sides of the inequality in (3·1) are random variables, and that w and $\hat{x}$ depend upon the a_i, as well as $u(a)$.

The method of solution is basically that due to Welch (1947*a*). It turns out to be possible to express $u(a)$ in the form

$$u(a)=u_0(a)+u_1(a)+u_2(a)+\ldots, \tag{3·2}$$

where each $u_t(a)$ is of order ν^{-t}, and where, if $\Sigma_{s=0}^{t} u_s(a)$ is written in place of $u(a)$ in (3·1), the equality is satisfied to within terms of order ν^{-t-1}. Large-sample theory gives the initial term $u_0(a)=\chi$, where

$$(2\pi)^{-\frac{1}{2}}\int_{-\chi}^{\chi} e^{-\frac{1}{2}t^2}\,dt=P. \tag{3·3}$$

The first corrective term can be derived by making the appropriate substitutions in the general theory presented in a previous paper (James, 1954, equation (2·53) or (2·59)). It is

$$u_1(a)=\tfrac{1}{4}(\chi^3-7\chi)\Sigma\frac{r_i^2}{\nu_i}+2\chi\Sigma\frac{r_i}{\nu_i}, \tag{3·4}$$

where $r_i=w_i/w$. The present problem is an example of what was there called the 'general case' of a linear hypothesis with variances requiring separate estimation; that is to say, the estimate $\hat{x}$ of μ is itself functionally dependent on the variance estimates. This makes for analytic complication, and only this first corrective term was given for the general problem considered there. (For the 'special case' where the estimates of the parameters in the linear hypothesis are functionally independent of the variance estimates a second correction term was given.) However, for the particular problem of weighted means the second and third corrective terms are manageable, although very lengthy calculations are involved. (Mrs Aspin recorded that the algebra for the fourth corrective term in Welch's v-problem ran to more than 100 pages; anyone wishing to tackle the same term in the present problem should be prepared for considerably more work than this.)

The following is a brief summary of Welch's computational technique. The left-hand side of (3·1) can be written as the average, over the distributions of the variance estimates a_i, of the conditional probability that $w^{\frac{1}{2}} \mid \hat{x}-\mu \mid \leqslant u(a)$ for *fixed* a_i. That is to say, our equation (3·1) for $u(a)$ can be written schematically as

$$\int \Pr[w^{\frac{1}{2}} \mid \hat{x}-\mu \mid \leqslant u(a) \mid a]\Pr[da]=P. \tag{3·5}$$

The left-hand side can be expressed as a series containing terms of order 1, ν^{-1}, ν^{-2}, ..., and depending upon $\Pr[\omega^{\frac{1}{2}} \mid \bar{x}-\mu \mid \leqslant u(\alpha)]$ (where $\bar{x}$ denotes $\Sigma\omega_i x_i/\omega$) and its derivatives with respect to the α_i. (The α_i involved parametrically in the *distribution* of $\bar{x}$ are regarded as constants in evaluating these derivatives, but not those involved in the *expressions* ω, $\bar{x}$ and

$u(\alpha)$.) The series (3·2) is now substituted for $u(\alpha)$ and terms of corresponding orders on both sides are equated. There results the following set of equations for $u_0(\alpha), u_1(\alpha), \ldots$, and hence for $u_0(a), u_1(a), \ldots$:

$$\Pr[\omega^{\frac{1}{2}} \mid \bar{x}-\mu \mid \leqslant u_0(\alpha)] = P, \quad \text{giving} \quad u_0(\alpha) = u_0(a) = \chi, \tag{3·6}$$

$$[u_1(\alpha)\, D + \Sigma\alpha_i^2 \partial_i^2/\nu_i] \Pr[\omega^{\frac{1}{2}} \mid \bar{x}-\mu \mid \leqslant \chi] = 0, \tag{3·7}$$

$$\begin{aligned}[u_2(\alpha)\, D + \tfrac{1}{2}u_1^2(\alpha)\, D^2 + (\Sigma\alpha_i^2\partial_i^2/\nu_i)\, u_1(\alpha)\, D \\ + \tfrac{4}{3}\Sigma\alpha_i^3\partial_i^3/\nu_i^2 + \tfrac{1}{2}\Sigma\alpha_i^2\alpha_j^2\partial_i^2\partial_j^2/\nu_i\nu_j] \Pr[\omega^{\frac{1}{2}} \mid \bar{x}-\mu \mid \leqslant \chi] = 0,\end{aligned} \tag{3·8}$$

$$\begin{aligned}[u_3(\alpha)\, D + u_2(\alpha)\, u_1(\alpha)\, D^2 + \tfrac{1}{6}u_1^3(\alpha)\, D^3 + (\Sigma\alpha_i^2\partial_i^2/\nu_i)\,(u_2(\alpha)\, D + \tfrac{1}{2}u_1^2(\alpha)\, D^2) \\ + (\tfrac{4}{3}\Sigma\alpha_i^3\partial_i^3/\nu_i^2 + \tfrac{1}{2}\Sigma\alpha_i^2\alpha_j^2\partial_i^2\partial_j^2/\nu_i\nu_j)\, u_1(\alpha)\, D + 2\Sigma\alpha_i^4\partial_i^4/\nu_i^3 \\ + \tfrac{4}{3}\Sigma\alpha_i^3\alpha_j^2\partial_i^3\partial_j^2/\nu_i^2\nu_j + \tfrac{1}{6}\Sigma\alpha_i^2\alpha_j^2\alpha_k^2\partial_i^2\partial_j^2\partial_k^2/\nu_i\nu_j\nu_k] \Pr[\omega^{\frac{1}{2}} \mid \bar{x}-\mu \mid \leqslant \chi] = 0.\end{aligned} \tag{3·9}$$

Here ∂_i denotes $\partial/\partial\alpha_i$ (interpreted in the sense indicated above) and D denotes $\partial/\partial\chi$. In a term like $\alpha_i^2\partial_i^2 u_1(\alpha)\, D \Pr[\ldots]$, ∂_i^2 operates on $u_1(\alpha)$ and on $\Pr[\ldots]$, but not on the initial α_i^2. The derivatives with respect to the α_i are most readily evaluated by the Taylor expansion technique used in my 1954 paper. After evaluating all the derivatives involved in (3·7), (3·8) and (3·9) and performing the summations we finally find the result

$$\begin{aligned}u(a) \quad \text{or} \quad u(r) = \chi[1 &+ \{\tfrac{1}{4}(\chi^2-7)\, U_{21} + 2U_{11}\} + \{\tfrac{1}{3}(\chi^4 - 31\chi^2 + 93)\, U_{32} \\ &+ \tfrac{1}{2}(11\chi^2 - 61)\, U_{22} + 8U_{12} + \tfrac{1}{32}(-9\chi^4 + 208\chi^2 - 543)\, U_{21}^2 \\ &+ \tfrac{3}{2}(-\chi^2+7)\, U_{21}U_{11} - 2U_{11}^2\} \\ &+ \{\tfrac{1}{8}(5\chi^6 - 417\chi^4 + 4957\chi^2 - 8997)\, U_{43} \\ &+ \tfrac{2}{3}(28\chi^4 - 643\chi^2 + 1668)\, U_{33} + 11(7\chi^2 - 33)\, U_{23} + 32U_{13} \\ &+ \tfrac{5}{12}(-3\chi^6 + 184\chi^4 - 1852\chi^2 + 2973)\, U_{32}U_{21} \\ &+ \tfrac{1}{3}\tfrac{0}{}(-\chi^4 + 31\chi^2 - 93)\, U_{32}U_{11} + \tfrac{1}{8}(-99\chi^4 + 1984\chi^2 - 4677)\, U_{22}U_{21} \\ &+ 3(-11\chi^2 + 61)\, U_{22}U_{11} + 6(-\chi^2+7)\, U_{12}U_{21} - 16U_{12}U_{11} \\ &+ \tfrac{1}{384}(243\chi^6 - 12053\chi^4 + 105569\chi^2 - 153063)\, U_{21}^3 \\ &+ \tfrac{7}{16}(9\chi^4 - 208\chi^2 + 543)\, U_{21}^3 U_{11} + \tfrac{15}{2}(\chi^2-7)\, U_{21}U_{11}^2 + 4U_{11}^3\}] + O(\nu^{-4}).\end{aligned} \tag{3·10}$$

where

$$U_{st} = \Sigma r_i^s/\nu_i^t, \quad r_i = w_i/w. \tag{3·11}$$

As a check on the lengthy algebra involved the calculation was repeated, but instead of starting with (3·1) the equivalent equation

$$\Pr[w(\hat{x}-\mu)^2 \leqslant 2h(a)] = P \tag{3·12}$$

for $2h(a) \equiv [u(a)]^2$ was used. The zero-order approximation to $h(a)$ is $\xi = \frac{1}{2}\chi^2$, where

$$\pi^{-\frac{1}{2}} \int_0^{\xi} t^{-\frac{1}{2}} e^{-t}\, dt = P, \tag{3·13}$$

and writing $h(a) = \xi + h_1(a) + \ldots$ we complete the solution as before. Finally, $u(a)$ is evaluated as the square root (to order ν^{-3}) of $2h(a)$.

4. TABLES FOR THE TWO-SAMPLE PROBLEM

In the case $k=2$, $u(r)$ is a function of the single quantity

$$r=r_1=\frac{w_1}{w_1+w_2}, \tag{4·1}$$

apart from ν_1, ν_2 and P.

The values for $(1-P)=0·10$ were worked out on an ordinary desk computer, using equation (3·10). These values were checked, and the values for $(1-P)=0·05$, $0·02$ and $0·01$ were worked out on the automatic digital computing machine at Manchester University.

The results have here been tabulated after rounding off to two decimal places. For infinite degrees of freedom tables of the normal distribution show that a rounding-off error of 0·005 implies a change in probability varying from about 0·002,1 when $(1-P)=0·10$ to about 0·000,29 when $(1-P)=0·01$; that is, to put the most unfavourable light on it, the maximum error due to this cause is less than 3 % of $(1-P)$ for the values tabulated. Since the values of $u(r)$ are always at least as large as the values with infinite degrees of freedom, rounding-off errors are likely to be even less serious in general. Moreover, when the degrees of freedom are only moderate a very considerable averaging effect takes place, for r varies from sample to sample and some tabular values are likely to be too small and others too large. For example, in an Appendix to Mrs Aspin's tables (1949) for the v-problem, Welch shows that, in the particular case $(1-P)=0·10$, $\nu_1=\nu_2=6$ (our notation) and equal population variances, the true probability using the two-place tables never differs from the nominal by more than about 0·000,4, compared with the bound 0·002,1 suggested above. I have not done any similar calculations using the two-place tabular values given here, but some details of the performance of the five-place figures from which they were derived are given in § 6.

In deciding on the minimum values of ν_1 and ν_2 to be given in the tables, the principle adopted was that alterations of at most two or three units shall be necessary in the second decimal place, should some improved method of calculation become available in the future. Of course guessing 'what the next term would be' is a dangerous procedure, but the third corrective term is never greater than 0·04; this value occurs when $(1-P)=0·01$, $\nu_1=\nu_2=10$, $r=0·3$ or $0·7$. A possible future change of several units in the last place has been accepted as a compromise between the desire to produce tables which are mathematically correct, and ones which shall be practically useful. The standard is lower than that imposed in the Welch–Aspin tables, where not only is the term of order ν^{-4} available for evaluation, but the series itself is more quickly convergent (in the practical sense of the word). If it were merely required that the probabilities should be close to their nominal values it would no doubt be possible to extend the range (cf. Table 6·1).

5. INTERPOLATION IN THE TABLES

For most purposes it will suffice to interpolate linearly with respect to all three arguments, although in certain cases this may lead to an error of several units in the second decimal place. In cases of doubt, harmonic interpolation should be used for the degrees of freedom. Harmonic interpolation should in any case be used for either ν_i in the panel which includes $\nu_i=\infty$.

6. The accuracy of the solution

The performance of the solution (3·10), or of any other proposed solution, may be checked by numerical integration. This process could no doubt also be made the basis of an iterative method of calculating $u(r)$, like that described by Trickett & Welch (1954) for the dual problem. The integral to be evaluated can be derived as follows.

The quantities $\chi_i^2 = \nu_i \omega_i / w_i$ are distributed independently of the x_i and of each other in χ^2-distributions with ν_i degrees of freedom. Write

$$\chi^2 = \Sigma \chi_i^2, \quad b_i = \chi_i^2 / \Sigma \chi_i^2 \quad (\Sigma b_i = 1). \tag{6·1}$$

Then χ^2 has a χ^2-distribution with $\nu_* = \Sigma \nu_i$ degrees of freedom, and is independent of the distribution of the b_i, which is

$$p(b)\,db = \frac{\Gamma(\frac{1}{2}\nu_*)}{\Pi\Gamma(\frac{1}{2}\nu_i)} \Pi b_i^{\frac{1}{2}\nu_i - 1}\, \Pi'\, db_i \quad (b_1, \ldots, b_k \geqslant 0;\ \Sigma b_i = 1). \tag{6·2}$$

Π denotes a product over $i = 1, \ldots, k$, and Π' a product over $i = 1, \ldots, (k-1)$ (say). We easily find

$$w^{\frac{1}{2}}(\hat{x} - \mu) = \frac{\Sigma w_i (x_i - \mu)}{w^{\frac{1}{2}}} = \frac{\Sigma \nu_i \omega_i (x_i - \mu)/b_i}{(\chi^2 \Sigma \nu_i \omega_i / b_i)^{\frac{1}{2}}}$$

$$= \left[\frac{\Sigma \nu_i \omega_i (x_i - \mu)/b_i}{(\Sigma \nu_i^2 \omega_i / b_i^2)^{\frac{1}{2}}} \left(\frac{\chi^2}{\nu_*}\right)^{-\frac{1}{2}}\right] \left[\frac{\Sigma \nu_i^2 \rho_i / b_i^2}{\nu_* \Sigma \nu_i \rho_i / b_i}\right]^{\frac{1}{2}}, \tag{6·3}$$

where

$$\rho_i = \omega_i / \omega, \quad \Sigma \rho_i = 1. \tag{6·4}$$

Now the conditional distribution of the first factor in (6·3), given the b_i, is that of t with ν_* degrees of freedom, by the usual rule. (Its marginal distribution is the same, but we make no use of this fact.) We also have

$$u(r) \quad \text{or} \quad u(r_i) = u\left(\frac{\nu_i \rho_i / b_i}{\Sigma \nu_i \rho_i / b_i}\right). \tag{6·5}$$

Therefore

$$\Pr[w^{\frac{1}{2}} \mid \hat{x} - \mu \mid \leqslant u(r)] = \int \Pr[w^{\frac{1}{2}} \mid \hat{x} - \mu \mid \leqslant u(r) \mid b]\, \Pr[db] \quad \text{(cf. (3·5))}$$

$$= \int F_{\nu_*}\left[\left(\frac{\nu_* \Sigma \nu_i \rho_i / b_i}{\Sigma \nu_i^2 \rho_i / b_i^2}\right)^{\frac{1}{2}} u\left(\frac{\nu_i \rho_i / b_i}{\Sigma \nu_i \rho_i / b_i}\right)\right] p(b)\, db, \tag{6·6}$$

where F denotes the ordinary t-integral:

$$F_\nu(t) = [B(\tfrac{1}{2}, \tfrac{1}{2}\nu)]^{-1} \int_{-t/\sqrt{\nu}}^{t/\sqrt{\nu}} (\tau^2 + 1)^{-\frac{1}{2}\nu - \frac{1}{2}}\, d\tau. \tag{6·7}$$

The integral in (6·6) is a $(k-1)$-fold one over the distribution (6·2). In the particular case $k = 2$ (6·6) may be written

$$\int_0^1 F_{\nu_*}\left[\left(\frac{\nu_* b b' (\nu_1 \rho b' + \nu_2 \rho' b)}{\nu_1^2 \rho b'^2 + \nu_2^2 \rho' b^2}\right)^{\frac{1}{2}} u\left(\frac{\nu_1 \rho b'}{\nu_1 \rho b' + \nu_2 \rho' b}\right)\right] \frac{b^{\frac{1}{2}\nu_1 - 1} b'^{\frac{1}{2}\nu_2 - 1}}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)}\, db, \tag{6·8}$$

where $\rho=\omega_1/\omega=(1-\rho')$ and $b'=(1-b)$. The integrals (6·6) and (6·8) simplify in various ways for equal degrees of freedom, and for $\rho_i \propto 1/\nu_i$. For example, when

$$k=2,\ \nu_1=\nu_2=\nu,\ \rho=\tfrac{1}{2},$$

(6·8) becomes

$$\int_0^1 F_{2\nu}\left[\left(\frac{2bb'}{b^2+b'^2}\right)^{\frac{1}{2}}u(b')\right]\frac{(bb')^{\frac{1}{2}\nu-1}}{B(\frac{1}{2}\nu,\ \frac{1}{2}\nu)}db. \tag{6·9}$$

The simplification in the argument of $u(r)$ is of great assistance in its numerical evaluation, for if a ten-panel integration formula is used no interpolation is necessary in the tables of $u(r)$, and interpolation is of the simplest kind with twenty panels.

The results recorded in Table 6·1 have been obtained in the case $k=2$, using the five-place values of $u(r)$ from which the two-place tables were compiled, and also the lower-order approximations correct to ν^{-2}, ν^{-1}, 1. The last-mentioned is the ordinary large-sample approximation.

Table 6·1. *Actual values of* $100\,(1-P)$ *compared with nominal values*

Order of approximation in the $1/\nu_i$	Nominal value of $(1-P)$				
	10% ($\nu_1=\nu_2=8$)			1% ($\nu_1=\nu_2=10$)	
	$\rho=0$ or 1	0·2 or 0·8	0·5	$\rho=0$ or 1	0·5
Zero	13·86	15·87	16·54	2·76	3·33
First	10·38	—	10·95	1·19	—
Second	10·03	—	10·07	1·02	—
Third	10·00	9·99	10·02	1·00	1·01

The calculations for the two-tailed 1 % point with $\nu_1=\nu_2=10$ and $\rho=0{\cdot}5$ were carried out because of the large value (about 0·04) of the third corrective term, $u_3(r)$, in the region $r=0{\cdot}3$–$0{\cdot}7$. It is seen that the actual probability is 1 % for practical purposes, even although the large-sample procedure would give rise to a probability over three times as great.

7. CONFIDENCE LIMITS FOR RATIOS

Finney (1950, 1952) has given some useful extensions of a method which was first given in its general form by Fieller (1940). These results all have to do with finding limits for ratios. Fieller's procedure is as follows. Suppose that y and z are normally distributed, unbiased estimates of η and ζ, and that their variances are $\lambda\sigma^2$ and $\lambda'\sigma^2$ and that their covariance is $\lambda''\sigma^2$; λ, λ', λ'' are known constants, but σ^2 is unknown and is independently estimated by s^2 based on ν degrees of freedom. It is required to find limits for $\mu=\eta/\zeta$ from the data. Now $(y-\mu z)/s(\lambda-2\mu\lambda''+\mu^2\lambda')^{\frac{1}{2}}$ has the t-distribution with ν degrees of freedom. Thus if t denotes the value only exceeded (numerically) with probability $(1-P)$, we have

$$\Pr[(y-\mu z)^2 \leqslant t^2 s^2(\lambda-2\mu\lambda''+\mu^2\lambda')]=P. \tag{7·1}$$

Thus a corresponding confidence region for μ consists of those values which satisfy the inequality on the left-hand side. The typical case is when $z^2-t^2s^2\lambda'>0$, that is to say when the denominator z is significantly different from zero (at level $(1-P)$). In this case the region is an ordinary interval, confidence limits for μ being

$$\mu_2,\mu_1=\frac{yz-t^2s^2\lambda''\pm ts[(z^2\lambda-2yz\lambda''+y^2\lambda')-t^2s^2(\lambda\lambda'-\lambda''^2)]^{\frac{1}{2}}}{z^2-t^2s^2\lambda'}$$

$$=\frac{m-g\lambda''/\lambda'\pm(ts/z)\,[(\lambda-2m\lambda''+m^2\lambda')-g(\lambda-\lambda''^2/\lambda')]^{\frac{1}{2}}}{1-g}, \tag{7·2}$$

where
$$m=y/z \quad \text{and} \quad g=t^2\lambda's^2/z^2. \tag{7·3}$$

But it is important to realize (because similar phenomena can occur in the generalizations considered below, although they will not be explicitly mentioned) that if $z^2-t^2s^2\lambda'<0$, i.e. if $g>1$, then the limits (7·2) become *exclusive*; that is to say, the region consists of the two parts $\mu\leqslant\mu_1$ and $\mu\geqslant\mu_2$ (taking $\mu_1<\mu_2$), and does not exclude the value $\mu=\infty$. It is even possible for the discriminant of the quadratic to become negative in this case, when the confidence region becomes $-\infty<\mu<\infty$; that is to say, the data are in agreement (at the given significance level) with any hypothetical value of μ. These phenomena are discussed in detail by Fieller (1954). The second form of (7·2) is useful because when g is very small it reduces to the result that would be obtained intuitively from the evaluation of the variance of m using 'statistical differentials' (and ignoring the inconvenient fact that this variance is really infinite).

Finney's generalizations make use of the Behrens–Fisher distribution, and lead to fiducial intervals, but not to confidence intervals in the direct probability sense. We now consider how these problems can be dealt with using the Welch–Aspin tables of v and the new tables of u.

First suppose that the conditions are the same as before, except that the variances of y and z are $\lambda\sigma^2$ and $\lambda'\sigma'^2$, while their covariance vanishes. Independent estimates s^2 and s'^2 of σ^2 and σ'^2, based on ν and ν' degrees of freedom, are available. Then $(y-\mu z)$ has zero expectation and variance $\lambda\sigma^2+\mu^2\lambda'\sigma'^2$ estimated by $\lambda s^2+\mu^2\lambda's'^2$. Thus

$$\Pr[(y-\mu z)^2\leqslant\{v(c)\}^2\,(\lambda s^2+\mu^2\lambda's'^2)]=P, \tag{7·4}$$

where
$$c=\lambda s^2/(\lambda s^2+\mu^2\lambda's'^2) \tag{7·5}$$

and $v(c)$ is the value given by the Welch–Aspin tables for ν and ν' degrees of freedom and probability P. In the case of a well-determined denominator (7·4) gives the confidence limits

$$\mu_2,\mu_1=\frac{m\pm(v/z)\,[(1-g)\,\lambda s^2+m^2\lambda's'^2]^{\frac{1}{2}}}{1-g}, \tag{7·6}$$

where $g=v^2\lambda's'^2/z^2$ and v is short for $v(c)$. Since c, and hence v, depends on μ_1 or μ_2, as the case may be, calculations have to be made iteratively using (7·5) and (7·6) alternately. As an initial approximation one could take $\mu=m$ in (7·5).

Now suppose that $\eta_1/\zeta_1=\ldots=\eta_k/\zeta_k=\mu$, and that we have unbiased and normally distributed estimates $y_1,\ldots,y_k,z_1,\ldots,z_k$ of $\eta_1,\ldots,\eta_k,\zeta_1,\ldots,\zeta_k$. The variances and covariance of y_i, z_i are $\lambda_i\sigma_i^2$, $\lambda_i'\sigma_i^2$, $\lambda_i''\sigma_i^2$, but different pairs are independent. Independently distributed estimates s_i^2 of the σ_i^2, based on ν_i degrees of freedom, are available. We require to find

confidence limits for μ. Now $x_i = y_i - \mu z_i$ has zero expectation and variance estimated by $(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2$. Therefore if we define

$$w_i = 1/(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2, \quad w = \Sigma w_i, \tag{7·7}$$

then

$$\Pr[(\Sigma w_i x_i)^2 \leqslant \{u(r)\}^2 w] = P, \tag{7·8}$$

where

$$r_i = \frac{w_i}{w} = \frac{1/(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2}{\Sigma[1/(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2]}, \tag{7·9}$$

and $u(r)$ is the value given by the tables in this paper (or their theoretical counterparts when $k > 2$) for the specified degrees of freedom and probability. Thus the corresponding confidence region is that of those values of μ which satisfy

$$\left[\Sigma \frac{y_i - \mu z_i}{(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2}\right]^2 \leqslant [u(r)]^2 \Sigma \frac{1}{(\lambda_i - 2\mu\lambda_i'' + \mu^2\lambda_i')\, s_i^2}. \tag{7·10}$$

This rather complicated inequality, which has to be considered in conjunction with the values of the r_i given by (7·9), may be of limited practical usefulness, although one imagines that when the denominators z_i are well determined it would give a single confidence interval, whose end-points could be found by an iterative procedure (cf. example 8·3). But a very simple case occurs when the constants λ_i, λ_i', λ_i'' are proportional, that is to say when

$$(\lambda_i, \lambda_i', \lambda_i'') = (\lambda, \lambda', \lambda'')\, \kappa_i, \quad \text{say.} \tag{7·11}$$

For example, the quantities y_i, z_i $(i = 1, \ldots, k)$ may have arisen from k different experiments, of fundamentally the same structure, but with different amounts of replication n_i. Then the constants κ_i are proportional to $1/n_i$. In this case (7·10) and (7·9) become simply

$$(\hat{y} - \mu\hat{z})^2 \leqslant [u(r)]^2 S^2(\lambda - 2\mu\lambda'' + \mu^2\lambda'), \tag{7·12}$$

where

$$r_i = \frac{1/\kappa_i s_i^2}{\Sigma(1/\kappa_i s_i^2)}, \quad \frac{1}{S^2} = \Sigma \frac{1}{\kappa_i s_i^2}, \tag{7·13}$$

and $\hat{y}, \hat{z}$ are weighted means calculated with weights $1/\kappa_i s_i^2$. (The estimates of the variances and covariance of $\hat{y}, \hat{z}$ are $(\lambda, \lambda', \lambda'')\, S^2$.) The r_i are now immediately calculable from the data, without any iteration, and in the case when $\hat{z}$ is significantly different from zero (7·12) gives at once the confidence limits

$$\mu_2, \mu_1 = \frac{m - g\lambda''/\lambda' \pm (uS/\hat{z})\,[(\lambda - 2m\lambda'' + m^2\lambda') - g(\lambda - \lambda''^2/\lambda')]^{\frac{1}{2}}}{1 - g}, \tag{7·14}$$

where

$$m = \hat{y}/\hat{z}, \quad g = u^2\lambda' S^2/\hat{z}^2, \tag{7·15}$$

and u is short for $u(r)$.

In his paper, Finney also considers the case where the weights used have been chosen without reference to the experimental data. This problem also admits a confidence-limit solution, but in this case we must use the Welch–Aspin v-tables instead of the u-tables. It is not necessary to choose the same weights for the y's and the z's, provided that it is possible to assume that $\eta_1 = \ldots = \eta_k$ and $\zeta_1 = \ldots = \zeta_k$. Thus suppose it has been decided to use weights W_i (independent of the s_i^2) for the y_i and W_i' for the z_i, with $\Sigma W_i = \Sigma W_i' = 1$. Let $\tilde{y}$ and $\tilde{z}$

denote the means $\Sigma W_i y_i$ and $\Sigma W'_i z_i$. Then the result is that the confidence limits for μ (for a significantly non-zero $\tilde{z}$) are

$$\mu_2, \mu_1 = \frac{m - gV''/V' \pm (v/\tilde{z})\,[(V - 2mV'' + m^2V') - g(V - V''^2/V')]^{\frac{1}{2}}}{1-g}, \tag{7·16}$$

where
$$m = \tilde{y}/\tilde{z}, \quad g = v^2V'/\tilde{z}^2, \quad v = v(c), \tag{7·17}$$

and
$$V = \Sigma W_i^2 \lambda_i s_i^2, \quad V' = \Sigma W_i'^2 \lambda'_i s_i^2, \quad V'' = \Sigma W_i W'_i \lambda''_i s_i^2, \tag{7·18}$$

and
$$c_i = \frac{(W_i^2\lambda_i - 2\mu W_i W'_i \lambda''_i + \mu^2 W_i'^2 \lambda'_i)\, s_i^2}{\Sigma(W_i^2\lambda_i - 2\mu W_i W'_i \lambda''_i + \mu^2 W_i'^2 \lambda'_i)\, s_i^2}. \tag{7·19}$$

Equations (7·16) and (7·19) have again to be solved iteratively, with $\mu = \mu_1$ or μ_2 in (7·19), as appropriate. This method seems to have no particular advantage over that using weights supplied by the data themselves, except when data from several independent sources are being combined, and the λ's are not proportional; in this case the simpler calculations might be a determining factor.

8. Examples of the use of the tables

Example 8·1. Twenty determinations of a quantity using a method of relatively low accuracy gave a mean result of 10·31 with a standard error (19 D.F.) of 1·08. A further set of ten determinations using a method of higher accuracy gave $11{\cdot}29 \pm 0{\cdot}67$. Assuming that both methods are unbiased, find 95 % confidence limits for the true value.

Since the standard errors already refer to the *means*, we have

$$w_1 = 1/(\lambda_1 s_1^2) = 1/(1{\cdot}08)^2 = 0{\cdot}857,$$
$$w_2 = 1/(\lambda_2 s_2^2) = 1/(0{\cdot}67)^2 = 2{\cdot}228.$$

Thus
$$\hat{x} = \frac{(0{\cdot}857)\,(10{\cdot}31) + (2{\cdot}228)\,(11{\cdot}29)}{0{\cdot}857 + 2{\cdot}228} = 11{\cdot}02,$$

and
$$r = \frac{0{\cdot}857}{0{\cdot}857 + 2{\cdot}228} = 0{\cdot}278.$$

Interpolation in Table 2 gives $u = 2{\cdot}29$. Thus the required 95 % confidence limits are

$$11{\cdot}02 \pm 2{\cdot}29/\sqrt{(0{\cdot}857 + 2{\cdot}228)} = 11{\cdot}02 \pm 1{\cdot}30.$$

Example 8·2. Finney (1952, § 9·3) quotes data of Gridgeman (1944), obtained in an assay of vitamin A, using growth rate in male rats as response. Thirty pairs (litter mates) of rats were used, ten pairs for each of three dose levels (0·9, 1·5, 2·5 units of vitamin A in the case of the standard preparation; 0·45, 0·75, 1·25 mg. of the test preparation). One of each pair of rats was assigned to the standard preparation and the other to the *corresponding* dose of the test preparation. If logarithms of dose (for convenience taken to the base $\frac{5}{3}$) are used as dose metameter, and growth in grams over 3 weeks as response metameter, then the dosage-response curves become nearly parallel straight lines, and the three dosage values are spaced one unit apart. The experiment can be analysed as a typical 'split-plot' design, 'whole plots' being pairs of rats and 'split plots' being individual rats. The mean difference in response between preparations (test minus standard) is $y = -5{\cdot}233$ gm. This is a comparison made *within* pairs of rats, for which the error mean square in the analysis of

variance is 39·35 (27 D.F.). The slope of the dosage-response lines is estimated as $z = 9{\cdot}350$ gm. per 5:3 increase in dose; but this is a contrast *between* pairs of rats, for which the error mean square is 98·60 (27 D.F.). According to the usual assumptions, y, z and the two error mean squares are independent. If the true potency of the test preparation is ρ units/mg., then $m = y/z = -0{\cdot}5597$ is an estimate of $\mu = \log_{5/3} 0{\cdot}5\rho$ (since the nominal potency of the test preparation, taken into account when fixing the doses to be administered, was 2·0 units/mg.). Thus the true potency is estimated as

$$R = 2(\tfrac{5}{3})^{-0{\cdot}5597} = 1{\cdot}50 \text{ units/mg.}$$

Now the estimated variances of y and z are $\lambda s^2 = 39{\cdot}35/15 = 2{\cdot}624$ and $\lambda' s'^2 = 98{\cdot}60/40 = 2{\cdot}465$. Taking $\mu = -0{\cdot}5597$ as a first approximation, equation (7·5) gives $c = 0{\cdot}77$. For 95 % limits, the first table in Trickett *et al.* (1956) gives $v = 2{\cdot}02$. Thus $g = 0{\cdot}115$, and equation (7·6) gives

$$\mu_1 = -1{\cdot}062, \quad \mu_2 = -0{\cdot}203.$$

Substituting these values back into (7·5) gives $c_1 = 0{\cdot}486$, $c_2 = 0{\cdot}963$, or $v_1 = 2{\cdot}00$, $v_2 = 2{\cdot}05$. These are only about 1 % different from the original values of v, and in practice it would probably not be worth while to carry the calculation further. However, substituting these values in (7·6), using v_1 with the minus sign and v_2 with the plus, we find the corrected limits

$$\mu_1 = -1{\cdot}055, \quad \mu_2 = -0{\cdot}198.$$

The corresponding 95 % confidence limits for the potency are 1·17 and 1·81 units/mg.

Example 8·3. Finney (1952, § 12·5) gives as another example data discussed by Bliss & Rose (1940) relating to an assay of parathyroid extract, the response being the serum calcium level in dogs injected with the standard and test preparations. The design is a fairly complicated one, based on a balanced incomplete block arrangement, and the reader is referred to the above authors for full details. Doses of 0·06 and 0·12 c.c./kg. body weight of each preparation were used, there being eighteen determinations at each of the four doses. For our purposes it will suffice to say that the analysis of variance splits into two independent parts, which will be referred to here as the interblock and intrablock analyses, each providing an estimate of potency and its own error mean square. The two parts can be regarded as separate experiments for further computational purposes. The first section gives, as an estimate of $\mu = \log_{\sqrt{2}} \rho$, the quantity y_1/z_1, where $y_1 = 4{\cdot}500$, $z_1 = 4{\cdot}833$, these having the estimated variances and covariance (30 D.F.) $\lambda_1 s_1^2 = 126{\cdot}8/6$, $\lambda_1' s_1^2 = 126{\cdot}8/24$, $\lambda_1'' s_1^2 = 0$. The second section gives the estimate y_2/z_2, where $y_2 = 3{\cdot}000$, $z_2 = 6{\cdot}417$, these having the estimated variances and covariance (30 D.F.) $\lambda_2 s_2^2 = 61{\cdot}35/12$, $\lambda_2' s_2^2 = 61{\cdot}35/48$, $\lambda_2'' s_2^2 = 0$. It will be seen that the λ's are proportional, and we may take, for example,

$$(\lambda_i, \lambda_i', \lambda_i'') = (\lambda, \lambda', \lambda'')\kappa_i = (\tfrac{1}{12}, \tfrac{1}{48}, 0)\kappa_i,$$

where $\kappa_1 = 2$, $\kappa_2 = 1$ (see equation (7·11)). The means, using weights $1/(\kappa_1 s_1^2)$, $1/(\kappa_2 s_2^2)$, are

$$\hat{y} = 3{\cdot}292, \quad \hat{z} = 6{\cdot}108.$$

Thus from (7·13) and (7·15),

$$m = 0{\cdot}5390, \quad r = 0{\cdot}195, \quad S^2 = 49{\cdot}40.$$

From Table 2, the appropriate value of u for 95 % limits is 2·07, so that $g = 0{\cdot}118$ and (7·14) gives the limits

$$\mu_1 = -0{\cdot}151, \quad \mu_2 = +1{\cdot}373.$$

The corresponding relative potencies are 0·949 and 1·61.

Up till now we have neglected a complication in Finney's analysis. This is the introduction of the body weight of the dogs as a concomitant variate in the intrablock analysis. (No significant reduction in error variance was obtained in the interblock analysis.) This yields the modified values

$$y_2 = 2{\cdot}511,\ z_2 = 6{\cdot}270,\ s_2^2 = 49{\cdot}74,\ \lambda_2 = 0{\cdot}083{,}93,\ \lambda_2' = 0{\cdot}020{,}89,\ \lambda_2'' = 0{\cdot}000{,}18.$$

The error degrees of freedom in this section of the analysis are now only 29. The introduction of the corrections for body weight have caused a slight correlation between y_2 and z_2, and the λ's are no longer quite proportional. However, equations (7·9) and (7·10) can be used; after some somewhat tedious arithmetic we find the limits 0·943, 1·53 for the relative potency ($P = 0{\cdot}95$).

9. Multivariate generalization of the weighted mean problem

The problem considered in § 1 can be generalized as follows. We suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are k p-variate vectors (for example, the means of k p-variate samples) which are distributed independently and normally about the same mean-point (centroid) $\boldsymbol{\xi}$, but with possibly different dispersion matrices $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$. The latter are unknown, but independent estimates of them (for example, the dispersion matrices of the k samples) are available; these are supposed to have Wishart distributions with $\nu_1, \ldots, \nu_k$ degrees of freedom. If $\mathbf{W}_i = \boldsymbol{\alpha}_i^{-1}$ and $\mathbf{W} = \Sigma \mathbf{W}_i$ and

$$\hat{\mathbf{x}} = \mathbf{W}^{-1} \Sigma \mathbf{W}_i \mathbf{x}_i, \tag{9·1}$$

then it is known that if the ν_i are large then $(\hat{\mathbf{x}} - \boldsymbol{\xi})' \mathbf{W} (\hat{\mathbf{x}} - \boldsymbol{\xi})$ has approximately a χ^2-distribution with p degrees of freedom. More precisely, making the appropriate substitutions in equation (5·28) of James (1954) we find that if we define

$$2h = \chi^2 + \Sigma \frac{1}{\nu_i} \left[2(p+1) \frac{\chi^2}{p} \operatorname{tr} \mathbf{W}^{-1} \mathbf{W}_i + \left(\frac{\chi^4}{p(p+2)} - 2\frac{\chi^2}{p} \right) \operatorname{tr} (\mathbf{W}^{-1} \mathbf{W}_i)^2 \right.$$
$$\left. + \frac{1}{2} \left(\frac{\chi^4}{p(p+2)} - 3\frac{\chi^2}{p} \right) (\operatorname{tr} \mathbf{W}^{-1} \mathbf{W}_i)^2 \right], \tag{9·2}$$

where

$$[\Gamma(\tfrac{1}{2}p)]^{-1} \int_0^{\frac{1}{2}\chi^2} t^{\frac{1}{2}p-1} e^{-t}\, dt = P, \tag{9·3}$$

then

$$\Pr[(\hat{\mathbf{x}} - \boldsymbol{\xi})' \mathbf{W} (\hat{\mathbf{x}} - \boldsymbol{\xi}) \leqslant 2h] = P + O(\nu^{-2}). \tag{9·4}$$

This equation can be used to give an ellipsoidal confidence region for the vector $\boldsymbol{\xi}$, with confidence coefficient approximately equal to P, if the degrees of freedom ν_i are reasonably large.

I would like to acknowledge the assistance of Dr D. W. J. Cruikshank, Miss D. E. Pilling and Mr J. F. P. Donovan, of the Department of Inorganic and Structural Chemistry, Leeds University, in facilitating the use of the Manchester University automatic digital computer (Mk. I).

REFERENCES

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika*, **35**, 88–96.

Aspin, A. A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, **36**, 290–6.

Behrens, W. V. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.* **68**, 807–37.

BLISS, C. I. & ROSE, C. L. (1940). The assay of parathyroid extract from the serum calcium of dogs. *Amer. J. Hyg.* A, **31**, 79–98.

CREASY, M. A. (1954). Limits for the ratio of means. *J. R. Statist. Soc.* B, **16**, 186–94.

FIELLER, E. C. (1940). The biological standardization of insulin. *J. R. Statist. Soc. Suppl.* **7**, 1–64.

FIELLER, E. C. (1954). Some problems in interval estimation. *J. R. Statist. Soc.* B, **16**, 175–85.

FINNEY, D. J. (1950). Two new uses of the Behrens–Fisher distribution. *J. R. Statist. Soc.* B, **12**, 293–300.

FINNEY, D. J. (1952). *Statistical Method in Biological Assay*. London: Griffin.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen., Lond.*, **6**, 391–8.

FISHER, R. A. (1939). The comparison of samples with possibly unequal variances. *Ann. Eugen., Lond.*, **9**, 174–80.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Ann. Eugen., Lond.*, **11**, 141–72.

FISHER, R. A. & YATES, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*. 4th. ed. Edinburgh: Oliver and Boyd.

GRIDGEMAN, N. T. (1944). *The Estimation of Vitamin A*. London: Lever Brothers and Unilever Ltd.

JAMES, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **41**, 19–43.

JEFFREYS, H. (1940). Note on the Behrens–Fisher formula. *Ann. Eugen., Lond.*, **10**, 48–51.

MAULDON, J. G. (1955). Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory. *J. R. Statist. Soc.* B, **17**, 79–85.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge: The University Press for the Biometrika Trustees.

SUKHATME, P. V. (1938). On Fisher and Behrens' test of significance for the difference in means of two normal samples. *Sankhyā*, **4**, 39–48.

TRICKETT, W. H. & WELCH, B. L. (1954). On the comparison of two means: further discussion of iterative methods for calculating tables. *Biometrika*, **41**, 361–74.

TRICKETT, W. H., WELCH, B. L. & JAMES, G. S. (1956). Further critical values for the two-means problem. *Biometrika*, **43**, 203–5.

WELCH, B. L. (1947*a*). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

WELCH, B. L. (1947*b*). On the Studentization of several variances. *Ann. Math. Statist.* **18**, 118–22.

YATES, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proc. Camb. Phil. Soc.* **35**, 579–91.

Table 1. *Upper 5 % critical values of* $u=(\hat{x}-\mu)\sqrt{(w_1+w_2)}$ *(i.e. upper 10 % critical values of* $|u|$*)*

	$r=\frac{w_1}{w_1+w_2}$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
ν_2	ν_1											
6	6	1·94	2·05	2·11	2·14	2·15	2·15	2·15	2·14	2·11	2·05	1·94
	8	1·94	2·01	2·05	2·07	2·08	2·07	2·07	2·06	2·03	1·96	1·86
	10	1·94	1·99	2·02	2·03	2·04	2·03	2·02	2·01	1·98	1·91	1·81
	15	1·94	1·97	1·98	1·99	1·99	1·98	1·97	1·95	1·91	1·85	1·75
	20	1·94	1·96	1·97	1·97	1·96	1·95	1·94	1·92	1·88	1·82	1·72
	∞	1·94	1·94	1·93	1·91	1·90	1·88	1·86	1·83	1·80	1·74	1·64
8	6	1·86	1·96	2·03	2·06	2·07	2·07	2·08	2·07	2·05	2·01	1·94
	8	1·86	1·93	1·97	1·99	2·00	2·00	2·00	1·99	1·97	1·93	1·86
	10	1·86	1·91	1·94	1·96	1·96	1·96	1·96	1·95	1·92	1·88	1·81
	15	1·86	1·89	1·90	1·91	1·91	1·91	1·90	1·89	1·86	1·82	1·75
	20	1·86	1·88	1·89	1·89	1·89	1·89	1·88	1·86	1·83	1·79	1·72
	∞	1·86	1·86	1·85	1·84	1·83	1·82	1·80	1·78	1·75	1·71	1·64
10	6	1·81	1·91	1·98	2·01	2·02	2·03	2·04	2·03	2·02	1·99	1·94
	8	1·81	1·88	1·92	1·95	1·96	1·96	1·96	1·96	1·94	1·91	1·86
	10	1·81	1·86	1·89	1·91	1·92	1·92	1·92	1·91	1·89	1·86	1·81
	15	1·81	1·84	1·86	1·87	1·87	1·87	1·87	1·85	1·83	1·80	1·75
	20	1·81	1·83	1·84	1·85	1·85	1·85	1·84	1·83	1·80	1·77	1·72
	∞	1·81	1·81	1·81	1·80	1·79	1·78	1·76	1·75	1·72	1·69	1·64
15	6	1·75	1·85	1·91	1·95	1·97	1·98	1·99	1·99	1·98	1·97	1·94
	8	1·75	1·82	1·86	1·89	1·90	1·91	1·91	1·91	1·90	1·89	1·86
	10	1·75	1·80	1·83	1·85	1·87	1·87	1·87	1·87	1·86	1·84	1·81
	15	1·75	1·78	1·80	1·81	1·82	1·82	1·82	1·81	1·80	1·78	1·75
	20	1·75	1·77	1·79	1·79	1·80	1·80	1·79	1·79	1·77	1·75	1·72
	∞	1·75	1·75	1·75	1·75	1·74	1·73	1·72	1·71	1·69	1·67	1·64
20	6	1·72	1·82	1·88	1·92	1·94	1·95	1·96	1·97	1·97	1·96	1·94
	8	1·72	1·79	1·83	1·86	1·88	1·89	1·89	1·89	1·89	1·88	1·86
	10	1·72	1·77	1·80	1·83	1·84	1·85	1·85	1·85	1·84	1·83	1·81
	15	1·72	1·75	1·77	1·79	1·79	1·80	1·80	1·79	1·79	1·77	1·75
	20	1·72	1·74	1·76	1·77	1·77	1·77	1·77	1·77	1·76	1·74	1·72
	∞	1·72	1·72	1·72	1·72	1·71	1·71	1·70	1·69	1·68	1·66	1·64
∞	6	1·64	1·74	1·80	1·83	1·86	1·88	1·90	1·91	1·93	1·94	1·94
	8	1·64	1·71	1·75	1·78	1·80	1·82	1·83	1·84	1·85	1·86	1·86
	10	1·64	1·69	1·72	1·75	1·76	1·78	1·79	1·80	1·81	1·81	1·81
	15	1·64	1·67	1·69	1·71	1·72	1·73	1·74	1·75	1·75	1·75	1·75
	20	1·64	1·66	1·68	1·69	1·70	1·71	1·71	1·72	1·72	1·72	1·72
	∞	1·64	1·64	1·64	1·64	1·64	1·64	1·64	1·64	1·64	1·64	1·64

$\hat{x}=(w_1x_1+w_2x_2)/(w_1+w_2)$ is the weighted mean of two independent normally distributed variables x_1 and x_2, which have the same expected value μ, and variances $\lambda_1\sigma_1^2$ and $\lambda_2\sigma_2^2$ respectively. Independently distributed estimates s_1^2, s_2^2 of σ_1^2, σ_2^2 based on ν_1, ν_2 degrees of freedom are available, and the weights used are $w_1=1/(\lambda_1 s_1^2)$, $w_2=1/(\lambda_2 s_2^2)$.

For example, if the x_i and s_i^2 $(i=1, 2)$ denote the means and variances of two samples of sizes n_i taken from two normal populations with the same mean μ, then $\nu_i=n_i-1$ and $\lambda_i=1/n_i$.

Table 2. *Upper 2½ % critical values of* $u=(\hat{x}-\mu)\sqrt{(w_1+w_2)}$ *(i.e. upper 5 % critical values of* $|u|$*)*

ν_2	$r=\frac{w_1}{w_1+w_2}$ / ν_1	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
8	8	2·31	2·37	2·43	2·45	2·47	2·47	2·47	2·45	2·43	2·37	2·31
	10	2·31	2·35	2·38	2·40	2·41	2·41	2·40	2·39	2·35	2·30	2·23
	12	2·31	2·34	2·36	2·37	2·37	2·37	2·36	2·34	2·31	2·25	2·18
	15	2·31	2·32	2·34	2·34	2·34	2·33	2·32	2·30	2·26	2·20	2·13
	20	2·31	2·31	2·32	2·31	2·31	2·30	2·28	2·26	2·22	2·16	2·09
	∞	2·31	2·29	2·27	2·24	2·22	2·20	2·17	2·14	2·10	2·04	1·96
10	8	2·23	2·30	2·35	2·39	2·40	2·41	2·41	2·40	2·38	2·35	2·31
	10	2·23	2·27	2·31	2·33	2·35	2·35	2·35	2·33	2·31	2·27	2·23
	12	2·23	2·26	2·29	2·30	2·31	2·31	2·30	2·29	2·27	2·23	2·18
	15	2·23	2·25	2·27	2·28	2·28	2·27	2·27	2·25	2·22	2·18	2·13
	20	2·23	2·24	2·25	2·25	2·25	2·24	2·23	2·21	2·18	2·14	2·09
	∞	2·23	2·21	2·20	2·18	2·16	2·14	2·12	2·09	2·06	2·02	1·96
12	8	2·18	2·25	2·31	2·34	2·36	2·37	2·37	2·37	2·36	2·34	2·31
	10	2·18	2·23	2·27	2·29	2·30	2·31	2·31	2·30	2·29	2·26	2·23
	12	2·18	2·21	2·24	2·26	2·27	2·27	2·27	2·26	2·24	2·21	2·18
	15	2·18	2·20	2·22	2·23	2·24	2·24	2·23	2·22	2·20	2·17	2·13
	20	2·18	2·19	2·20	2·21	2·21	2·20	2·19	2·18	2·16	2·13	2·09
	∞	2·18	2·17	2·16	2·14	2·13	2·11	2·09	2·07	2·04	2·00	1·96
15	8	2·13	2·20	2·26	2·30	2·32	2·33	2·34	2·34	2·34	2·32	2·31
	10	2·13	2·18	2·22	2·25	2·27	2·27	2·28	2·28	2·27	2·25	2·23
	12	2·13	2·17	2·20	2·22	2·23	2·24	2·24	2·23	2·22	2·20	2·18
	15	2·13	2·16	2·18	2·19	2·20	2·20	2·20	2·19	2·18	2·16	2·13
	20	2·13	2·15	2·16	2·17	2·17	2·17	2·16	2·15	2·14	2·11	2·09
	∞	2·13	2·12	2·11	2·10	2·09	2·08	2·06	2·04	2·02	1·99	1·96
20	8	2·09	2·16	2·22	2·26	2·28	2·30	2·31	2·31	2·32	2·31	2·31
	10	2·09	2·14	2·18	2·21	2·23	2·24	2·25	2·25	2·25	2·24	2·23
	12	2·09	2·13	2·16	2·18	2·19	2·20	2·21	2·21	2·20	2·19	2·18
	15	2·09	2·11	2·14	2·15	2·16	2·17	2·17	2·17	2·16	2·15	2·13
	20	2·09	2·10	2·12	2·13	2·13	2·14	2·13	2·13	2·12	2·10	2·09
	∞	2·09	2·08	2·07	2·07	2·06	2·05	2·03	2·02	2·00	1·98	1·96
∞	8	1·96	2·04	2·10	2·14	2·17	2·20	2·22	2·24	2·27	2·29	2·31
	10	1·96	2·02	2·06	2·09	2·12	2·14	2·16	2·18	2·20	2·21	2·23
	12	1·96	2·00	2·04	2·07	2·09	2·11	2·13	2·14	2·16	2·17	2·18
	15	1·96	1·99	2·02	2·04	2·06	2·08	2·09	2·10	2·11	2·12	2·13
	20	1·96	1·98	2·00	2·02	2·03	2·05	2·06	2·07	2·07	2·08	2·09
	∞	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96

$\hat{x}=(w_1x_1+w_2x_2)/(w_1+w_2)$ is the weighted mean of two independent normally distributed variables x_1 and x_2, which have the same expected value μ, and variances $\lambda_1\sigma_1^2$ and $\lambda_2\sigma_2^2$ respectively. Indedependently distributed estimates s_1^2, s_2^2 of σ_1^2, σ_2^2 based on ν_1, ν_2 degrees of freedom are available, and the weights used are $w_1=1/(\lambda_1 s_1^2)$, $w_2=1/(\lambda_2 s_2^2)$.

For example, if the x_i and s_i^2 $(i=1, 2)$ denote the means and variances of two samples of sizes n_i taken from two normal populations with the same mean μ, then $\nu_i=n_i-1$ and $\lambda_i=1/n_i$.

Table 3. *Upper* 1 % *critical values of* $u=(\hat{x}-\mu)\sqrt{(w_1+w_2)}$ (*i.e. upper* 2 % *critical values of* $|u|$)

	$r=\frac{w_1}{w_1+w_2}$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
ν_2	ν_1											
10	10	2·76	2·80	2·84	2·87	2·89	2·89	2·89	2·87	2·84	2·80	2·76
	12	2·76	2·79	2·81	2·83	2·83	2·83	2·82	2·80	2·77	2·73	2·68
	15	2·76	2·77	2·78	2·78	2·78	2·78	2·76	2·74	2·70	2·65	2·60
	20	2·76	2·76	2·75	2·75	2·74	2·73	2·71	2·68	2·64	2·58	2·53
	30	2·76	2·75	2·73	2·71	2·70	2·68	2·65	2·62	2·57	2·52	2·46
	∞	2·76	2·73	2·69	2·66	2·63	2·59	2·55	2·51	2·46	2·40	2·33
12	10	2·68	2·73	2·77	2·80	2·82	2·83	2·83	2·83	2·81	2·79	2·76
	12	2·68	2·71	2·74	2·76	2·77	2·77	2·77	2·76	2·74	2·71	2·68
	15	2·68	2·69	2·71	2·72	2·72	2·72	2·71	2·70	2·67	2·64	2·60
	20	2·68	2·68	2·68	2·68	2·68	2·67	2·66	2·64	2·61	2·57	2·53
	30	2·68	2·67	2·66	2·65	2·64	2·62	2·60	2·58	2·54	2·50	2·46
	∞	2·68	2·65	2·62	2·60	2·57	2·54	2·51	2·47	2·43	2·38	2·33
15	10	2·60	2·65	2·70	2·74	2·76	2·78	2·78	2·78	2·78	2·77	2·76
	12	2·60	2·64	2·67	2·70	2·71	2·72	2·72	2·72	2·71	2·69	2·68
	15	2·60	2·62	2·64	2·66	2·66	2·67	2·66	2·66	2·64	2·62	2·60
	20	2·60	2·61	2·62	2·62	2·62	2·62	2·61	2·60	2·58	2·55	2·53
	30	2·60	2·60	2·59	2·59	2·58	2·57	2·56	2·54	2·52	2·49	2·46
	∞	2·60	2·58	2·56	2·54	2·51	2·49	2·46	2·44	2·40	2·37	2·33
20	10	2·53	2·58	2·64	2·68	2·71	2·73	2·74	2·75	2·75	2·76	2·76
	12	2·53	2·57	2·61	2·64	2·66	2·67	2·68	2·68	2·68	2·68	2·68
	15	2·53	2·55	2·58	2·60	2·61	2·62	2·62	2·62	2·62	2·61	2·60
	20	2·53	2·54	2·55	2·56	2·57	2·57	2·57	2·56	2·55	2·54	2·53
	30	2·53	2·53	2·53	2·53	2·53	2·53	2·52	2·51	2·49	2·48	2·46
	∞	2·53	2·51	2·50	2·48	2·46	2·44	2·43	2·40	2·38	2·35	2·33
30	10	2·46	2·52	2·57	2·62	2·65	2·68	2·70	2·71	2·73	2·75	2·76
	12	2·46	2·50	2·54	2·58	2·60	2·62	2·64	2·65	2·66	2·67	2·68
	15	2·46	2·49	2·52	2·54	2·56	2·57	2·58	2·59	2·59	2·60	2·60
	20	2·46	2·48	2·49	2·51	2·52	2·53	2·53	2·53	2·53	2·53	2·53
	30	2·46	2·46	2·47	2·48	2·48	2·48	2·48	2·48	2·47	2·46	2·46
	∞	2·46	2·45	2·44	2·43	2·41	2·40	2·39	2·38	2·36	2·34	2·33
∞	10	2·33	2·40	2·46	2·51	2·55	2·59	2·63	2·66	2·69	2·73	2·76
	12	2·33	2·38	2·43	2·47	2·51	2·54	2·57	2·60	2·62	2·65	2·68
	15	2·33	2·37	2·40	2·44	2·46	2·49	2·51	2·54	2·56	2·58	2·60
	20	2·33	2·35	2·38	2·40	2·43	2·44	2·46	2·48	2·50	2·51	2·53
	30	2·33	2·34	2·36	2·38	2·39	2·40	2·41	2·43	2·44	2·45	2·46
	∞	2·33	2·33	2·33	2·33	2·33	2·33	2·33	2·33	2·33	2·33	2·33

$\hat{x}=(w_1x_1+w_2x_2)/(w_1+w_2)$ is the weighted mean of two independent normally distributed variables x_1 and x_2, which have the same expected value μ, and variances $\lambda_1\sigma_1^2$ and $\lambda_2\sigma_2^2$ respectively. Independently distributed estimates s_1^2, s_2^2 of σ_1^2, σ_2^2 based on ν_1, ν_2 degrees of freedom are available, and the weights used are $w_1=1/(\lambda_1 s_1^2)$, $w_2=1/(\lambda_2 s_2^2)$.

For example, if the x_i and s_i^2 $(i=1, 2)$ denote the means and variances of two samples of sizes n_i taken from two normal populations with the same mean μ, then $\nu_i=n_i-1$ and $\lambda_i=1/n_i$.

Table 4. *Upper ½ % critical values of* $u=(\hat{x}-\mu)\sqrt{(w_1+w_2)}$ *(i.e. upper 1 % critical values of* $|u|$*)*

	$r=\frac{w_1}{w_1+w_2}$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
ν_2	ν_1											
10	10	3·17	3·20	3·24	3·27	3·29	3·30	3·29	3·27	3·24	3·20	3·17
	12	3·17	3·18	3·20	3·21	3·22	3·22	3·21	3·18	3·14	3·09	3·05
	15	3·17	3·16	3·16	3·16	3·16	3·15	3·13	3·10	3·05	3·00	2·95
	20	3·17	3·14	3·13	3·11	3·10	3·08	3·05	3·02	2·96	2·90	2·85
	30	3·17	3·13	3·10	3·07	3·05	3·02	2·99	2·94	2·88	2·82	2·75
	∞	3·17	3·11	3·06	3·01	2·96	2·91	2·86	2·80	2·74	2·66	2·58
12	10	3·05	3·09	3·14	3·18	3·21	3·22	3·22	3·21	3·20	3·18	3·17
	12	3·05	3·07	3·10	3·12	3·14	3·14	3·14	3·12	3·10	3·07	3·05
	15	3·05	3·06	3·07	3·07	3·08	3·07	3·06	3·04	3·01	2·98	2·95
	20	3·05	3·04	3·03	3·03	3·02	3·01	2·99	2·96	2·93	2·88	2·85
	30	3·05	3·03	3·01	2·99	2·97	2·95	2·92	2·89	2·85	2·80	2·75
	∞	3·05	3·01	2·97	2·93	2·89	2·84	2·80	2·75	2·70	2·64	2·58
15	10	2·95	3·00	3·05	3·10	3·13	3·15	3·16	3·16	3·16	3·16	3·17
	12	2·95	2·98	3·01	3·04	3·06	3·07	3·08	3·07	3·07	3·06	3·05
	15	2·95	2·96	2·98	2·99	3·00	3·00	3·00	2·99	2·98	2·96	2·95
	20	2·95	2·94	2·95	2·95	2·95	2·94	2·93	2·91	2·89	2·87	2·85
	30	2·95	2·93	2·92	2·91	2·90	2·88	2·86	2·84	2·81	2·78	2·75
	∞	2·95	2·91	2·88	2·85	2·81	2·78	2·75	2·71	2·67	2·62	2·58
20	10	2·85	2·90	2·96	3·02	3·05	3·08	3·10	3·11	3·13	3·14	3·17
	12	2·85	2·88	2·93	2·96	2·99	3·01	3·02	3·03	3·03	3·04	3·05
	15	2·85	2·87	2·89	2·91	2·93	2·94	2·95	2·95	2·95	2·94	2·95
	20	2·85	2·85	2·86	2·87	2·88	2·88	2·88	2·87	2·86	2·85	2·85
	30	2·85	2·84	2·84	2·83	2·83	2·82	2·81	2·80	2·78	2·77	2·75
	∞	2·85	2·82	2·80	2·77	2·75	2·72	2·70	2·67	2·64	2·61	2·58
30	10	2·75	2·82	2·88	2·94	2·99	3·02	3·05	3·07	3·10	3·13	3·17
	12	2·75	2·80	2·85	2·89	2·92	2·95	2·97	2·99	3·01	3·03	3·05
	15	2·75	2·78	2·81	2·84	2·86	2·88	2·90	2·91	2·92	2·93	2·95
	20	2·75	2·77	2·78	2·80	2·81	2·82	2·83	2·83	2·84	2·84	2·85
	30	2·75	2·75	2·76	2·76	2·76	2·77	2·76	2·76	2·76	2·75	2·75
	∞	2·75	2·73	2·72	2·70	2·69	2·67	2·65	2·63	2·62	2·60	2·58
∞	10	2·58	2·66	2·74	2·80	2·86	2·91	2·96	3·01	3·06	3·11	3·17
	12	2·58	2·64	2·70	2·75	2·80	2·84	2·89	2·93	2·97	3·01	3·05
	15	2·58	2·62	2·67	2·71	2·75	2·78	2·81	2·85	2·88	2·91	2·95
	20	2·58	2·61	2·64	2·67	2·70	2·72	2·75	2·77	2·80	2·82	2·85
	30	2·58	2·60	2·62	2·63	2·65	2·67	2·69	2·70	2·72	2·73	2·75
	∞	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58

$\hat{x}=(w_1x_1+w_2x_2)/(w_1+w_2)$ is the weighted mean of two independent normally distributed variables x_1 and x_2, which have the same expected value μ, and variances $\lambda_1\sigma_1^2$ and $\lambda_2\sigma_2^2$ respectively. Indedependently distributed estimates s_1^2, s_2^2 of σ_1^2, σ_2^2 based on ν_1, ν_2 degrees of freedom are available, and the weights used are $w_1=1/(\lambda_1 s_1^2)$, $w_2=1/(\lambda_2 s_2^2)$.

For example, if the x_i and s_i^2 $(i=1, 2)$ denote the means and variances of two samples of sizes n_i taken from two normal populations with the same mean μ, then $\nu_i=n_i-1$ and $\lambda_i=1/n_i$.

ON ESTIMATING THE LATENT AND INFECTIOUS PERIODS OF MEASLES

II. FAMILIES WITH THREE OR MORE SUSCEPTIBLES

By NORMAN T. J. BAILEY

Design and Analysis of Scientific Experiment, 6 Keble Road, Oxford

1. Introduction

In a previous paper (Bailey (1956), referred to here as Part I) I discussed, for families with only two susceptibles, the maximum-likelihood estimation of parameters in an epidemic model involving a normally distributed latent period after the receipt of infection, followed by a constant infectious period terminating with the appearance of symptoms and removal of the patient from circulation. The present paper gives the extensions required for dealing with families having more than two susceptibles. The case of three susceptibles, which is quite common, is described in some detail as it presents a few new features. Larger families can be analysed in a similar way, but the procedure becomes rapidly more complicated, especially as the misclassification of links in the chain develops into an item of major importance. Since there is at present little data on such families we shall give an indication only of how to take account of all the complications involved.

An illustrative example is provided here for families with three susceptibles using some excellent material that has very kindly been made available to me by Dr R. E. Hope Simpson. The data are based on families residing in the Cirencester area during 1946–52 with three susceptible children under 15 years of age including at least one case of measles.

2. Description of data for families with three susceptibles

As shown in Part I, for families with only two susceptibles, the data, consisting of say a total of N families with at least one case, fall naturally into three parts. There are A families with two cases, both having been infected simultaneously by an outside contact. There are a further B families with two cases, where the second case is derived from the first by cross-infection within the family. We also know the time interval between the two cases, w for the first type and z for the second. As a first approximation we assume that these two types, which are labelled (2) and (1^2) in chain-binomial notation (Bailey, 1955), can be accurately distinguished. The third type of family, of which there are C, so that $N = A+B+C$, contains only a single case and is labelled (1).

When dealing with families having three susceptibles the parts of the data involving one or two cases only can again be described as above, and we shall use the same notation. But we also have in addition D families containing three cases, so that now $N = A+B+C+D$. These families are of four kinds represented by the chain-binomial symbols (1^3), (12), (3) and (21), with actual numbers E, F, G and H, respectively, where $D = E+F+G+H$. This time there are two time intervals to be recorded: u between the first and second cases, and v between the second and third cases. Assuming for the time being that the different types of chain can be correctly identified, we must consider the type of information they provide in some detail.

First, take the E families of type (1^3), where the second case is derived from the first, and the third from the second, both by cross-infection within the family. The two intervals, u and v, are clearly each a 'z-type' variable such as arises from the B families giving a (1^2) pattern. Secondly, the F families of type (12), where the second and third cases have been simultaneously infected by the first, also provides two z-type observations, but these are now given by u and $u+v$. We therefore have in all $J = B+2(E+F)$ z-type observations which can be subjected to the analysis already presented in Part I.

Next, we consider the G families of type (3), where all three cases have been simultaneously infected by an outside contact. These must be examined separately as they cannot be put in terms of other types of data already discussed. The appropriate analysis is given in the next section.

The H families of type (21), where the first two cases are derived from a simultaneous infection and the third by cross-infection, present a special difficulty. The first interval u is plainly a 'w-type' variable and can be taken in conjunction with the data for the A families of type (2), giving $K = A+H$ families altogether. The second interval v has a very much more complicated distribution, since the third case could have been infected during either of the infectious periods of the first two cases. It seems better to ignore the small amount of information available from this source in order to avoid excessive complexity.

Table 1. *Distribution of time interval for fifteen families of three susceptibles with two cases*

Time interval in days	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No. of families	.	2	.	1	1	.	2	1	1	.	.	3	1	2	1
Probable type of chain	(2)					(1^2)									
Total no.	$A=4$					$B=11$									

Finally, we can analyse the data according to the numbers of families showing different types of chain, apart from the time intervals involved. The main subdivision here depends on whether the disease is introduced by a single case or by two simultaneously infected cases. A triple introduction adds nothing further to what has already been mentioned in the last paragraph but one. When there are two initial cases, i.e. A of type (2) or H of type (21), the treatment is very similar to that given in Part I for the distribution of B, given $B+C$. When there is a single introduction we have types (1), (1^2), (1^3) and (12), with observed numbers C, B, E and F, where we write $M = B+C+E+F$.

A new point arises here in connexion with the possibility of variations in the chance of infection. As shown in Part I the z-type distribution is relatively insensitive to such variations, while the w-type variables are quite unaffected. The mean frequencies of A and H will also be unchanged. Variation in the chance of infection is therefore ignored so far as those sections of the data are concerned. However, it has been shown (Bailey, 1953) that the chain-binomial analysis of families of three susceptibles with a single initial case *is* considerably influenced by variations in the chance of infection. An additional parameter must therefore be introduced when this part of the data is analysed.

The last matter to be mentioned in this section is the question of correctly identifying the links of the chains involved. The distinction between (1^2) and (2) is accomplished in a similar manner to that described in Part I. Table 1 shows the distribution of fifteen observations

of this sort with an approximate dichotomy being made, as before, between 4 and 5 days. It is a little more difficult to separate the four groups of families with three cases and we now need to inspect a two-way table. The distribution of fifty-seven observations is shown in

Table 2. *Distribution of time intervals* (u, v) *for fifty-seven families of three susceptibles with three cases*

Time (v) between second and third cases in days	Time (u) between first and second cases in days																Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
0	.	1	1	.	.	.	.	.	.	6	1	3	1	1	.	.	.	14
1	2	1	.	.	.	.	1	.	.	.	3	3	1	.	.	.	1	12
2	1	.	.	.	.	.	1	1	.	2	2	3	.	.	2	.	.	12
3	.	.	1	.	.	.	.	.	1	1	.	1	.	.	.	.	.	4
4	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	2
5	.	.	.	.	.	.	1	.	2	.	.	.	.	.	.	.	.	3
6	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1
7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1
9	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2
10	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
11	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
12	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1
14	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Total	6	4	3	1	.	.	3	3	3	9	6	10	2	2	3	1	1	$57 = D$

Table 2*a*. *Summary of Table* 2 *giving probable types of chain with observed numbers, based on a dividing line between* 4 *and* 5 *days*

(3) $G = 7$	(12) $F = 37$
(21) $H = 7$	(1^3) $E = 6$

Tables 2 and 2*a*. Inspection of the marginal totals suggests that an appropriate division can again be made between 4 and 5 days, but there is now a somewhat greater possibility of overlap than appeared in Part I. As mentioned in § 5 below the chance of misclassification can be allowed for by the introduction of additional parameters, but seems hardly worth

embarking on in the present investigation. Tables 3 and 4 give the amalgamated expected and observed numbers for those parts of the data giving rise to w- and z-type distributions, while analyses of the chains for families with two and three cases are presented in Table 5.

Table 3. *Amalgamated frequencies for w-distribution, with a total of $K = A + H$ observations*

Time interval in days	Observed no.		Expected no.	
0	3	7	1·87	5·29
1	4		3·42	
2	1	4	2·61	5·71
3	2		1·64	
4	1		0·88	
⩾5	.		0·58	
Total	11 = K		11·00	

Table 4. *Amalgamated frequencies for z-distribution, with a total of $J = B + 2(E + F)$ observations*

Time interval in days	Observed no.		Expected no.	
0	.	8	.	4·32
1	.		.	
2	.		0·01	
3	.		0·04	
4	.		0·24	
5	3		1·01	
6	5		3·02	
7	5		6·64	
8	6		10·97	
9	17		14·16	
10	7	31	15·00	28·80
11	24		13·80	
12	10		11·55	
13	10		8·87	
14	5		6·03	
15	1	5	3·43	5·66
16	3		1·54	
17	1		0·53	
⩾18	.		0·16	
Total	97 = J		97·00	

Table 5. *Analysis of chains for two or three cases*

Type of introduction	Type of chain	Observed no.	Expected no.
Single	(1)	6	8·62
	(1^2)	11	6·23
	(1^3)	6	9·81
	(12)	37	35·34
	Total	$60 = M$	60·00
Double	(2)	4	3·05
	(21)	7	7·95
	Total	$11 = K$	11·00

3. Derivation of scores and information functions

For a detailed discussion of the mathematical model being used reference should be made to Part I. It is sufficient to repeat here that we assume the latent period, x, to be normally distributed with mean m and variance σ^2, and the ensuing infectious period to be of constant length a. Infection of further susceptibles during this time is taken to be a Poisson process such that the chance of a given susceptible contracting the disease in time dt is $\lambda\,dt$. The five main aspects of the data that are faily easily analysed will now be described in detail.

The w- and z-type distributions arising from the data are easily dealt with according to the theory already described in Part I. The only difference being that there all contributions to scores and information functions were amalgamated (see (9) and (11) of Part I), whereas here it is more convenient to have the individual contributions displayed separately.

w-type distribution

There are K families giving w-type observations, and the appropriate score and information functions are

$$S_\sigma = K\sigma^{-1}(\tfrac{1}{2}V\sigma^{-2} - 1), \tag{1}$$

$$I_{\sigma\sigma} = 2K\sigma^{-2}, \tag{2}$$

where, as before, V is the observed second moment of the distribution about the origin.

z-type distribution

Here we have J families in all yielding z-type observations. The four scores are

$$\left.\begin{aligned} S_\lambda &= J\{m - \bar{z} + \lambda^{-1} + \lambda\sigma^2 - a(e^{\lambda a} - 1)^{-1}\} + \sum_z T_\lambda, \\ S_a &= -J\lambda(e^{\lambda a} - 1)^{-1} + \sum_z T_a, \\ S_m &= J\lambda + \sum_z T_m, \\ S_\sigma &= J\lambda^2\sigma + \sum_z T_\sigma, \end{aligned}\right\} \tag{3}$$

where $\bar{z}$ and v are the mean and variance of the observed distribution and T_θ is defined as in Part I and summed over the observed values of z. The corresponding information functions are

$$\left.\begin{aligned}
I_{\lambda\lambda} &= \sum_z T_\lambda^2 \quad -J\{\lambda^2\sigma^4+\sigma^2-\lambda^{-2}+a^2e^{\lambda a}(e^{\lambda a}-1)^{-2}\},\\
I_{\lambda a} &= \sum_z T_\lambda T_a +J\{(1+\lambda^2\sigma^2)\,(e^{\lambda a}-1)^{-1}-\lambda a e^{\lambda a}(e^{\lambda a}-1)^{-2}\},\\
I_{\lambda m} &= \sum_z T_\lambda T_m -J(1+\lambda^2\sigma^2),\\
I_{\lambda\sigma} &= \sum_z T_\lambda T_\sigma -J\lambda^3\sigma^3,\\
I_{aa} &= \sum_z T_a^2 \quad -J\lambda^2(e^{\lambda a}-1)^{-2},\\
I_{am} &= \sum_z T_a T_m +J\lambda^2(e^{\lambda a}-1)^{-1},\\
I_{a\sigma} &= \sum_z T_a T_\sigma +J\lambda^3\sigma(e^{\lambda a}-1)^{-1},\\
I_{mm} &= \sum_z T_m^2 \quad -J\lambda^2,\\
I_{m\sigma} &= \sum_z T_m T_\sigma -J\lambda^3\sigma,\\
I_{\sigma\sigma} &= \sum_z T_\sigma^2 \quad -J\lambda^4\sigma.
\end{aligned}\right\} \tag{4}$$

Chains with double introductions

We now take the K families with a double introduction, i.e. A of type (2) and H of type (21). Using the Greenwood rather that the Reed–Frost formulation (see Bailey, 1955, §§ 6 and 7), the relative frequencies are $e^{-\lambda a}$ and $1-e^{-\lambda a}$. Contributions to the scores and information functions are therefore

$$S_\lambda = Ha(e^{\lambda a}-1)^{-1}-Aa, \quad S_a = \lambda a^{-1}S_\lambda, \tag{5}$$

$$I_{\lambda\lambda} = Ka^2(e^{\lambda a}-1)^{-1}, \quad I_{\lambda a} = \lambda a^{-1}I_{\lambda\lambda}, \quad I_{aa} = \lambda^2 a^{-2}I_{\lambda\lambda}. \tag{6}$$

There is no special difficulty in developing corresponding formulae for the Reed–Frost variant if required, though variation in the chance of infection must then be allowed for.

Chains with single introductions

As mentioned above, it is necessary to take into account the possibility of variations in the chance of infection when analysing the frequencies of the several chains starting with only a single introductory case. This has already been discussed at length in an earlier paper (Bailey, 1953). The chief assumption made was that the chance of cross-infection, p, which when constant equals $1-e^{-\lambda a}$ in our present notation, followed a β-distribution with parameters x and y, namely,

$$dF = \frac{1}{B(x,y)}p^{x-1}(1-p)^{y-1}\,dp \quad (0\leqslant p\leqslant 1). \tag{7}$$

The scores and observed information functions, repeated here for convenience using the present symbols for observed quantities, were shown to be

$$\left.\begin{aligned} S'_x &= \left\{\frac{B+E+F}{x}+\frac{E+F}{x+1}\right\}-\left\{\frac{M}{x+y}+\frac{M}{x+y+1}+\frac{B+E}{x+y+2}\right\}, \\ S'_y &= \left\{\frac{B+C+E}{y}+\frac{B+C}{y+1}\right\}-\left\{\frac{M}{x+y}+\frac{M}{x+y+1}+\frac{B+E}{x+y+2}\right\}, \end{aligned}\right\} \tag{8}$$

$$\left.\begin{aligned} I'_{xx} &= \frac{B+E+F}{x^2}+\frac{E+F}{(x+1)^2}-W, \\ I'_{xy} &= -W, \\ I'_{yy} &= \frac{B+C+E}{y^2}+\frac{B+C}{(y+1)^2}-W, \end{aligned}\right\} \tag{9}$$

where

$$W=\frac{M}{(x+y)^2}+\frac{M}{(x+y+1)^2}+\frac{B+E}{(x+y+2)^2}.$$

Primes are used here to distinguish these *auxiliary* scores and information functions from those with which we are more immediately concerned.

This procedure entails using two parameters to specify the probability of infection. The parameter λ, which we are already using to score data that are relatively insensitive to variations in the chance of infection, can be regarded as an average probability. We therefore need to introduce one additional parameter. The average value of p for the distribution (7) is $\bar{p}=x/(x+y)$, so that we can write

$$\frac{x}{x+y}=1-e^{-\lambda a},$$

or

$$x=y(e^{\lambda a}-1). \tag{10}$$

We can now suppose the expected frequencies to have been written in terms of λ, a and y instead of x and y. The usual processes of differentiation, using the functional relation (10), then lead to the following expressions for the scores and information functions as required in the present context, written in terms of those given in (8) and (9):

$$S_\lambda = a(x+y)S'_x, \quad S_a=\lambda(x+y)S'_x, \quad S_y=\left(\frac{x}{y}\right)S'_x+S'_y, \tag{11}$$

$$\left.\begin{aligned} I_{\lambda\lambda} &= a^2(x+y)\{(x+y)I'_{xx}-S'_x\}, \\ I_{\lambda a} &= \lambda a^{-1}I_{\lambda\lambda}-(x+y)S'_x, \\ I_{\lambda y} &= a(x+y)y^{-1}(xI'_{xx}+yI'_{xy}-S'_x), \\ I_{aa} &= \lambda^2 a^{-2}I_{\lambda\lambda}, \\ I_{ay} &= \lambda a^{-1}I_{\lambda y}, \\ I_{yy} &= \left(\frac{x}{y}\right)^2 I'_{xx}+2\left(\frac{x}{y}\right)I'_{xy}+I'_{yy}. \end{aligned}\right\} \tag{12}$$

Time-interval distribution for triple introductions

The last item to be discussed is the extraction of information from the G families involving a triple introduction. Each basic observation is the number-pair (u, v), where $u \leqslant v$. Let the times at which symptoms occur in the three patients, measuring from the common

point of infection as origin, be ξ_1, ξ_2 and ξ_3. Then these variables are normally and independently distributed with mean $m+a$ and variance σ^2. The joint-frequency distribution of the *ordered* trio (ξ_1, ξ_2, ξ_3) is

$$f(\xi_1 \leqslant \xi_2 \leqslant \xi_3)\, d\xi_1 d\xi_2 d\xi_3 = \frac{3!}{(2\pi\sigma^2)^{\frac{3}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{3}(\xi_i - m - a)^2\right\} d\xi_1 d\xi_2 d\xi_3. \tag{13}$$

If we now use the transformation

$$\left.\begin{aligned} u &= \xi_2 - \xi_1, \\ v &= \xi_3 - \xi_2, \\ s &= \xi_1 + \xi_2 + \xi_3, \end{aligned}\right\} \tag{14}$$

to write down the joint distribution of u, v and s, we can then integrate out s to give the required joint distribution of u and v only. This turns out to be

$$f(u, v)\, du dv = \frac{\sqrt{3}}{\pi\sigma^2} \exp\left\{-\frac{1}{3\sigma^2}(u^2 + uv + v^2)\right\} du dv. \tag{15}$$

We can now derive a score and information function for σ in the usual way. If for a set of observed number-pairs (u_i, v_i) $(i = 1, \ldots, G)$, we put

$$Z = \sum_{i=1}^{G}(u_i^2 + u_i v_i + v_i^2), \tag{16}$$

we obtain

$$S_\sigma = -2G\sigma^{-1} + \tfrac{2}{3}Z\sigma^{-3}, \quad I_{\sigma\sigma} = 4G\sigma^{-2}. \tag{17}$$

These results are, moreover, not affected by variations in the chance of infection.

This now completes the derivation of all the various components of the scores and information functions required for obtaining maximum-likelihood estimates of the five parameters λ, a, m, σ and y. The only elements of the 5×5 information matrix which have not been mentioned explicitly are $I_{m\lambda}$ and $I_{\sigma\lambda}$, and these are easily seen to be identically zero. In analysing actual data we first classify them according to the five main aspects above. Next, we calculate in each case, for trial values of the parameters, the appropriate contributions to scores and information functions. Corresponding contributions are then added and the resultant vector of overall scores multiplied by the complete information matrix to give first corrections to the trial values. The process is then repeated as usual until the desired accuracy is achieved.

4. ILLUSTRATIVE EXAMPLE

The foregoing maximum-likelihood scores and information functions can now be applied to Hope Simpson's data, the relevant aspects of which are exhibited in Tables 1–5. A preliminary inspection of the data shows that so far as the z-distribution in Table 4 goes a really satisfactory fit is unlikely to be obtained on account of the apparent excess of observations at 11 days with a counterbalancing deficit at 10 days. In Part I a difficulty of this sort was encountered with families of two in the appearance of possible spurious peaks at 7 and 14 days. This was thought to be due to an unconscious bias associated with integral multiples of a week. In the present case no very satisfactory explanation can be discerned. However, as the agreement between hypothesis and observation seems in other respects to be quite good, the best plan seems to be to pool the frequencies for 10 and 11 days when proceeding to test goodness-of-fit.

Preliminary estimates for λ, a, m and σ were taken from the final values previously obtained for families of two, while a trial value of y was available from an earlier analysis of data from Providence, Rhode Island (Bailey, 1953). After carrying out the standard procedure of maximum-likelihood scoring, the final estimates turned out to be

$$\left.\begin{aligned} \hat{\lambda} &= 0{\cdot}180 \pm 0{\cdot}039, \\ \hat{a} &= 7{\cdot}05 \pm 1{\cdot}13 \text{ days}, \\ \hat{m} &= 7{\cdot}63 \pm 0{\cdot}50 \text{ day}, \\ \hat{\sigma} &= 1{\cdot}59 \pm 0{\cdot}26 \text{ day}, \\ \hat{y} &= 0{\cdot}56 \pm 0{\cdot}32. \end{aligned}\right\} \tag{18}$$

These estimates of λ, a, m and σ are rather less precise than those given in equation (21) of Part I for families of two, but it may be noted that in no case are the two estimates of any parameter significantly different. The parameter y is not determined with much precision, although the estimate obtained does suggest an appreciable amount of variation in the chance of infection from family to family.

In testing goodness-of-fit, based as in Part I on the last set of estimates but one in the iteration, the classes bracketed together have been pooled to avoid small expectations. We have also amalgamated the frequencies for 10 and 11 days in the z-distribution as indicated above. The total number of degrees of freedom is 13, i.e. 1 from the w-distribution in Table 3, 8 from the z-distribution in Table 4 and 4 in all from Table 5, while 5 must be removed to allow for the parameters estimated. We actually obtain an overall χ^2 of 14·6 on 8 D.F. As the 5 % point is at 15·5 we can regard the fit as reasonably satisfactory, except of course for the anomalous behaviour of the frequencies in the 10- and 11-day classes of the z-distribution. Whether this is due to some bias in collecting the data, or whether it is of genuine biological significance is a matter which requires further investigation, and should be given special attention when new data of this type are collected.

Some consideration should be given at this point to the consequences of neglecting the effect of variations in λ on the form of the z-distribution. Using the method of § 5 in Part I, it can be shown that the fairly substantial variation envisaged there would in the present case actually improve the fit very slightly, reducing χ^2 by about 0·5.

5. Extension to larger families

The general procedure described here for families of three can clearly be extended to larger families. By picking out contributions to w- and z-type distributions we could use the scores and information functions given in (1)–(4) directly. When analysing the different kinds of chain allowing for a variable chance of infection the formulae given above in (10)–(12) would have to be applied to the extensions of (8) and (9) indicated in detail in Bailey (1953). Double and triple introductions can also be dealt with as above. Other multiple introductions require the obvious extensions of (13)–(17). We have in the present paper neglected to make use of the distribution of v-intervals in chains of type (21), because of undue complexity. With larger families there would be further relatively intractable items of this sort. Another point of importance is that with small families the errors introduced by neglecting to make allowance for the probability of chains being misclassified are likely to be small. But in larger families this source of error would be much more pronounced because of the greater opportunity for the distributions of different kinds of chains to overlap. If sufficient data

of this kind are forthcoming the difficulty could be tackled along the lines suggested in § 6 of Part I.

It is again a pleasure to acknowledge my indebtedness to Dr R. E. Hope Simpson, of the Cirencester Public Health Laboratory Service, for making available to me the epidemiological records used for the illustrative example in § 4, and to express my thanks to Mrs Tamara Hazlewood for undertaking the computations required in obtaining the numerical results.

REFERENCES

BAILEY, N. T. J. (1953). The use of chain-binomials with a variable chance of infection for the analysis of intra-household epidemics. *Biometrika*, **40**, 279.

BAILEY, N. T. J. (1955). Some problems in the statistical analysis of epidemic data. *J. R. Statist. Soc.* B. **17**, 35.

BAILEY, N. T. J. (1956). On estimating the latent and infectious periods of measles. I. Families with two susceptibles only. *Biometrika*, **43**, 1.

SIGNIFICANCE TESTS FOR A VARIABLE CHANCE OF INFECTION IN CHAIN-BINOMIAL THEORY

By NORMAN T. J. BAILEY

Design and Analysis of Scientific Experiment, 6 Keble Road, Oxford

In their analysis of measles data for Providence, Rhode Island; Wilson, Bennett, Allen & Worcester (1939) found that, although Greenwood's chain-binomial model (Greenwood, 1931) fitted satisfactorily the distribution of the total number of cases in families of given size, this theory was inadequate when the data were analysed according to the different types of chains involved. Greenwood (1949) suggested that this might be due to variations in the chance of infection, p, between different households. Subsequently it was shown by Bailey (1953) that good agreement between theory and observation could be obtained on the assumption that p varied according to the β-distribution

$$dF = \frac{1}{B(x,y)} p^{x-1}(1-p)^{y-1}\,dp \quad (0 \leqslant p \leqslant 1). \tag{1}$$

The appropriate scores and information functions were given for estimating the two parameters, x and y, in households of three and four susceptibles.

Now if in fact p is constant, x and y are both infinite and the maximum-likelihood method breaks down. Dr Armitage pointed this out in the discussion on my Royal Statistical Society paper (Bailey, 1955), and suggested that one might find it preferable to work with the reciprocals $x' = x^{-1}$ and $y' = y^{-1}$. This is certainly more satisfactory, but the maximum-likelihood scoring technique again fails in the limit as x' and y' tend to zero, since the scores contain terms proportional to x'^{-1} and y'^{-1}. We cannot derive an adequate significance test for variation in the chance of infection without making some assumption about the limiting form of the ratios y/x and y'/x'. It follows from (1) that the average chance of infection, $\bar{p}$, is given by

$$\bar{p} = \frac{x}{x+y} = \left(1+\frac{y}{x}\right)^{-1}. \tag{2}$$

One way out of the difficulty is therefore to use $\bar{p}$ as one parameter, and x' or y' as the other. The only drawback here is the algebraic complexity of the scores and information functions. An alternative and much simpler formulation is obtained by writing

$$p = \frac{x}{x+y}, \quad z = \frac{1}{x+y}, \tag{3}$$

where, for convenience, we have dropped the bar from $\bar{p}$, and $p = 1-q$.

Table 1 gives the expected frequencies of different types of chains in families of three for the Greenwood model, the modified model involving x and y, and the alternative form of the latter using p and z. Observations from the Providence measles data are also included. Scores for p and z are

$$\left.\begin{aligned} S_p &= \frac{b+c+d}{p} - \frac{a+b+c}{q} + \frac{c+d}{p+z} - \frac{a+b}{q+z}, \\ S_z &= -\frac{n}{1+z} - \frac{2(b+c)}{1+2z} + \frac{c+d}{p+z} + \frac{a+b}{q+z}, \end{aligned}\right\} \tag{4}$$

while the observed information functions are

$$\left.\begin{aligned} I_{pp} &= \frac{b+c+d}{p^2} + \frac{a+b+c}{q^2} + \frac{c+d}{(p+z)^2} + \frac{a+b}{(q+z)^2}, \\ I_{pz} &= \qquad\qquad \frac{c+d}{(p+z)^2} - \frac{a+b}{(q+z)^2}, \\ I_{zz} &= -\frac{n}{(1+z)^2} - \frac{4(b+c)}{(1+2z)^2} + \frac{c+d}{(p+z)^2} + \frac{a+b}{(q+z)^2}. \end{aligned}\right\} \tag{5}$$

The expressions appearing in (4) and (5) are particularly suited to rapid computation, since for trial values of p and z the various quotients involved in each line can be accumulated directly on a calculating machine, the squares of denominators in the case of information functions being read directly from Barlow's tables.

Table 1. *Expected and observed numbers of chains in families of three susceptibles*

Type of chain	Expected nos. on Greenwood model	Expected nos. on modified model		Observed nos.	Providence measles data
		In terms of x and y	In terms of p and z		
1	nq^2	$\frac{ny(y+1)}{(x+y)(x+y+1)}$	$\frac{nq(q+z)}{(1+z)}$	a	34
1^2	$2npq^2$	$\frac{2nxy(y+1)}{(x+y)(x+y+1)(x+y+2)}$	$\frac{2npq(q+z)}{(1+z)(1+2z)}$	b	25
1^3	$2np^2q$	$\frac{2nx(x+1)y}{(x+y)(x+y+1)(x+y+2)}$	$\frac{2npq(p+z)}{(1+z)(1+2z)}$	c	36
12	np^2	$\frac{nx(x+1)}{(x+y)(x+y+1)}$	$\frac{np(p+z)}{(1+z)}$	d	239
Total	n	n	n	n	334

Suppose, however, we are merely concerned to test for the existence of variability in the chance of infection, i.e. to examine the hypothesis that $z=0$. Then we can avoid the full scoring procedure involved in estimating p and z jointly by basing the significance test on the distribution of the score S_z, calculated at the values $z=0$ and $p=\hat{p}_0$, where $\hat{p}_0$ is the maximum-likelihood estimate of p when $z=0$. Thus

$$\hat{p}_0 = \frac{b+2c+2d}{2a+3b+3c+2d}, \tag{6}$$

and

$$S_z(\hat{p}_0, 0) = \frac{c+d}{\hat{p}_0} + \frac{a+b}{1-\hat{p}_0} - (a+3b+3c+d). \tag{7}$$

The large sample variance of $S_z(\hat{p}_0, 0)$ is easily shown to be $[I_{zz} - I_{pz}^2/I_{pp}]_{z=0}$. Using expected information functions we therefore have

$$\operatorname{var} S_z(\hat{p}_0, 0) = \frac{n(3-7pq)}{1+pq}. \tag{8}$$

The data shown in Table 1 give $\hat{p}_0 = 0{\cdot}789$ and $S_z = 172 \pm 23$. This result is strongly significant, as expected from the earlier investigation (Bailey, 1953).

Similar results are easily obtained for larger families. Table 2 gives the expectations and observations appropriate to families of four. The frequencies expected in terms of x and y have been omitted for reasons of space, but have been set out before in Table 4 of my 1953 paper. This time the scores for p and z are

$$\left.\begin{aligned} S_p &= \frac{n-a}{p} - \frac{n-h}{q} + \frac{n-a-b}{p+z} + \frac{e+f+g+h}{p+2z} - \frac{n-g-h}{q+z} - \frac{a+b+c+e}{q+2z} - \frac{b+c}{q+3z}, \\ S_z &= -\frac{n}{1+z} - \frac{2n}{1+2z} - \frac{3(n-a-h)}{1+3z} - \frac{4(b+c+e+f)}{1+4z} - \frac{5(c+e)}{1+5z} \\ &\quad + \frac{n-a-b}{p+z} + \frac{2(e+f+g+h)}{p+2z} + \frac{n-g-h}{q+z} + \frac{2(a+b+c+e)}{q+2z} + \frac{3(b+c)}{q+3z}, \end{aligned}\right\} \tag{9}$$

with information functions

$$\left.\begin{aligned} I_{pp} &= \frac{n-a}{p^2} + \frac{n-h}{q^2} + \frac{n-a-b}{(p+z)^2} + \frac{e+f+g+h}{(p+2z)^2} + \frac{n-g-h}{(q+z)^2} + \frac{a+b+c+e}{(q+2z)^2} + \frac{b+c}{(q+3z)^2}, \\ I_{pz} &= \frac{n-a-b}{(p+z)^2} + \frac{2(e+f+g+h)}{(p+2z)^2} - \frac{n-g-h}{(q+z)^2} - \frac{2(a+b+c+e)}{(q+2z)^2} - \frac{3(b+c)}{(q+3z)^2}, \\ I_{zz} &= -\frac{n}{(1+z)^2} - \frac{4n}{(1+2z)^2} - \frac{9(n-a-h)}{(1+3z)^2} - \frac{16(b+c+e+f)}{(1+4z)^2} - \frac{25(c+e)}{(1+5z)^2} \\ &\quad + \frac{n-a-b}{(p+z)^2} + \frac{4(e+f+g+h)}{(p+2z)^2} + \frac{n-g-h}{(q+z)^2} + \frac{4(a+b+c+e)}{(q+2z)^2} + \frac{9(b+c)}{(q+3z)^2}. \end{aligned}\right\} \tag{10}$$

Again we see that the scores and information functions are in a form suitable for easy computation.

The short significance test for variation in infectiousness is now given by calculating

$$S_z(\hat{p}_0, 0) = \frac{(c+d)+3(e+f+g+h)}{\hat{p}_0} + \frac{3(a+e)+6(b+c)+(d+f)}{1-\hat{p}_0} - (3a+10b+15c+6d+15e+10f+6g+3h), \tag{11}$$

where

$$\hat{p}_0 = \frac{b+2(c+d)+3(e+f+g+h)}{3a+5b+6c+4d+6e+5f+4g+3h}. \tag{12}$$

This time the expression for the variance of the score does not appear to reduce to a convenient simple formula, but we can still avoid the full iterative procedure by using, as before,

$$\operatorname{var} S_z(\hat{p}_0, 0) = [I_{zz} - I_{pz}^2/I_{pp}]_{z=0}, \tag{13}$$

but where, for $z = 0$, we now substitute the *observed* information functions given by

$$\left.\begin{aligned} I_{pp} &= \frac{b+2(c+d)+3(e+f+g+h)}{\hat{p}_0^2} + \frac{3(a+e)+4(b+c)+2(d+f)+g}{(1-\hat{p}_0)^2}, \\ I_{pz} &= \frac{(c+d)+3(e+f+g+h)}{\hat{p}_0^2} - \frac{3(a+e)+6(b+c)+(d+f)}{(1-\hat{p}_0)^2}, \\ I_{zz} &= \frac{(c+d)+5(e+f+g+h)}{\hat{p}_0^2} + \frac{5(a+e)+14(b+c)+(d+f)}{(1-\hat{p}_0)^2} \\ &\quad - \{5(a+h)+30(b+f)+55(c+e)+14(d+g)\}. \end{aligned}\right\} \tag{14}$$

If the Reed–Frost or Lidwell variant of the chain-binomial model (see Bailey, 1953, 1955) were though necessary, certain obvious modifications of the above would be required.

Applications of these formulae to the data in Table 2 gives $\hat{p}_0 = 0{\cdot}791$ and $S_z = 157 \pm 41$, again a highly significant result.

The modification discussed above can also be applied to a recent development of chain-binomial theory (Bailey, 1956) in which account is taken of variations in the incubation period and infectiousness is not confined to an instant but is allowed to persist for a fixed

Table 2. *Expected and observed numbers in families of four susceptibles*

Type of chain	Expected nos. on Greenwood model	Expected nos. on modified model	Observed nos.	Providence measles data
1	nq^3	$\frac{nq(q+z)\,(q+2z)}{(1+z)\,(1+2z)}$	a	4
1^2	$3npq^4$	$\frac{3npq(q+z)\,(q+2z)\,(q+3z)}{(1+z)\,(1+2z)\,(1+3z)\,(1+4z)}$	b	3
1^3	$6np^2q^4$	$\frac{6npq(p+z)\,(q+z)\,(q+2z)\,(q+3z)}{(1+z)\,(1+2z)\,(1+3z)\,(1+4z)\,(1+5z)}$	c	1
12	$3np^2q^2$	$\frac{3npq(p+z)\,(q+z)}{(1+z)\,(1+2z)\,(1+3z)}$	d	8
1^4	$6np^3q^3$	$\frac{6npq(p+z)\,(p+2z)\,(q+z)\,(q+2z)}{(1+z)\,(1+2z)\,(1+3z)\,(1+4z)\,(1+5z)}$	e	4
1^22	$3np^3q^2$	$\frac{3npq(p+z)\,(p+2z)\,(q+z)}{(1+z)\,(1+2z)\,(1+3z)\,(1+4z)}$	f	3
121	$3np^3q$	$\frac{3npq(p+z)\,(p+2z)}{(1+z)\,(1+2z)\,(1+3z)}$	g	10
13	np^3	$\frac{np(p+z)\,(p+2z)}{(1+z)\,(1+2z)}$	h	67
Total	n	n	n	100

interval of time. This model uses four basic parameters. In families of two the data are insensitive to variations in infectiousness between families, but with families of three, however, this must be allowed for in analysing that part of the data which yields chains of the types shown in Table 1. One additional parameter was therefore introduced—the quantity called y here. Difficulties again arise if we wish to test whether $y' = y^{-1}$ is significantly different from zero. The solution is to use z for the additional parameter, as above. We therefore have *auxiliary* scores and information functions S'_p, S'_z; I'_{pp}, I'_{pz} and I'_{zz}, given by (4) and (5) above with primes added, corresponding to (8) and (9) of the 1956 paper. The relation $p = 1 - e^{-\lambda a}$ is then used to derive the contributions to scores and information functions for the parameters λ, a and z. We thus have

$$S_\lambda = aqS'_p, \quad S_a = \lambda q S'_p, \quad S_z = S'_z, \tag{15}$$

and

$$\left.\begin{aligned} I_{\lambda\lambda} &= a^2q(qI'_{pp}+S'_p), & I_{\lambda a} &= \lambda a^{-1}I_{\lambda\lambda}-qS'_p,\\ I_{\lambda z} &= aqI'_{pz}, & I_{aa} &= \lambda^2a^{-2}I_{\lambda\lambda},\\ I_{az} &= \lambda a^{-1}I_{\lambda z}, & I_{zz} &= I'_{zz}, \end{aligned}\right\} \tag{16}$$

corresponding to (11) and (12) of the 1956 paper. We could find maximum-likelihood estimates of the four parameters λ, a, m and σ, given $z=0$, and then test the significance of z by comparing S_z with its standard error. However, it is probably easier in practice first to test z on the chained part of the data only, using expressions (6), (7) and (8) above. We then proceed to estimate all five parameters, using as a trial value of z, either zero if the preliminary test is not significant, or the actual estimate from this part of the data, based on (4) and (5), if it is.

REFERENCES

Bailey, N. T. J. (1953). The use of chain-binomials with a variable chance of infection for the analysis of intra-household epidemics. *Biometrika*, **40**, 279.

Bailey, N. T. J. (1955). Some problems in the statistical analysis of epidemic data. *J. R. Statist. Soc.* B, **17**, 35.

Bailey, N. T. J. (1956). On estimating the latent and infectious periods of measles. II. Families with three or more susceptibles. *Biometrika*, **43**, 322.

Greenwood, M. (1931). On the statistical measure of infectiousness. *J. Hyg., Camb.*, **31**, 336.

Greenwood, M. (1949). The infectiousness of measles. *Biometrika*, **36**, 1.

Wilson, E. B., Bennett, C., Allen, M. & Worcester, J. (1939). Measles and scarlet fever in Providence, R.I., 1929–34 with respect to age and size of family. *Proc. Amer. Phil. Soc.*, **80**, 357.

ON THE VARIATION OF YIELD VARIANCE WITH PLOT SIZE

BY P. WHITTLE

Applied Mathematics Laboratory, New Zealand Department of Scientific and Industrial Research, Wellington

The problem examined is that of evaluating the spatial covariance function of yield density, from a knowledge of the way yield variance varies with plot size and shape. Results are obtained in § 3 for several kinds of plot. Results are also obtained (§ 4) on the dependence of the yield variance on plot geometry for very small and very large plots. Special attention is paid to the case for which the covariance follows a power law at large distances.

1. INTRODUCTION

It is well known that, in order to explain the observed variation of yield variance with size and shape of plot, it is necessary to allow the possibility of correlation between yield densities at any two points in the plot. (We shall restrict ourselves to the stationary case, for which the expected yield density is constant over the area.) Moreover, it appears that this spatial correlation must often fall off relatively slowly with increasing distance between the two points; as a power function of the distance rather than as an exponential.

The same behaviour is shown by observations on yarn diameter, flood height (Feller, 1951), and response from population samples. The calculations of this article will apply to these cases, too, but for concreteness we shall continue to speak of plots and yields (although we shall sometimes use the word 'region' instead of 'plot', indicating that we do not confine ourselves to two dimensions).

The type of calculation most usually made is to evaluate the yield variance for a plot of definite size and shape, and for a given spatial covariance function $\rho(s)$. However, the inverse calculation would probably be more useful in general: to determine the covariance function from a knowledge of yield variance as a function of plot geometry. Such a procedure would enable one to make use of experimental results to obtain at least a partial estimate of $\rho(s)$. We consider this question in § 3.

A solution of one form of problem is most easily reached by using a Mellin transform, so that the covariance which falls off as a power of the distance,

$$\rho(s) \sim C\,|\,s\,|^{-\nu} \quad (s \text{ large}), \tag{1}$$

makes an unforced début.

There is a good deal of evidence to show that many observed covariance functions are of the form (1) (Fairfield Smith, 1938). This is all the more interesting, in view of the fact that the simpler linear models of yield variation do not predict a power law, but rather a much more quickly diminishing law of the type

$$\rho(s) \sim C\,|\,s\,|^{-\nu} e^{-\alpha|s|} \quad (s \text{ large}). \tag{2}$$

We shall pursue this question further in § 4.

It is easy to show that the 'yield' variance of a large one-dimensional interval is asymptotically proportional to a power of the length of the interval if the covariance function is of the form (1), but analogous results are not obtained quite so immediately in two or more dimensions. This topic is also covered in §§ 3 and 4.

2. Formulae for the yield variance

We shall use a vector co-ordinate $\mathbf{x}$ to denote points in the plane. Let us consider a plot, or region, Ω, with area A and yield $y(\Omega)$. (We use the words 'plane' and 'area', but most formulae of the paper apply to regions in n-dimensional space, $n = 1, 2, 3, \ldots$.)

Consider two regions Ω_1 and Ω_2 which contract to zero in such a way that ultimately they constitute two points a distance $\mathbf{s}$ apart. We shall define the covariance function as

$$\rho(\mathbf{s}) = \lim_{A_1, A_2 \to 0} \frac{\operatorname{cov}[y(\Omega_1), y(\Omega_2)]}{A_1 A_2}, \tag{3}$$

and shall restrict ourselves to cases for which this limit exists.

If $V(\Omega)$ is the population variance of $y(\Omega)$, then it follows from definition (3) that

$$V(\Omega) = \int_\Omega \int_\Omega \rho(\mathbf{x}_1 - \mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2. \tag{4}$$

If $\rho(\mathbf{s})$ is continuous at the origin then we can draw one immediate conclusion from (4):

$$V(\Omega) \sim \rho(\mathbf{0})\, A^2 \quad (\Omega \text{ small}), \tag{5}$$

that is, for small regions (i.e. for regions which are small in all their dimensions), V is proportional to the square of the area. This is to be contrasted with the case of zero spatial correlation, when V is proportional to area under all circumstances.

For the one-dimensional case, when Ω consists of a segment of a line of length a, then (4) simplifies to

$$V(a) = 2 \int_0^a (a - x)\, \rho(x)\, dx. \tag{6}$$

For the case when Ω consists of a rectangle in a plane, with sides a, b parallel to the co-ordinate axes, we have similarly

$$V(a, b) = 4 \int_0^b \int_0^a (a - x)\, (b - y)\, \rho(x, y)\, dx\, dy. \tag{7}$$

The integral (6) may be evaluated for the commoner choices of $\rho(s)$, but (7) and other two-dimensional formulae can generally be evaluated only for rather unlikely forms of the function $\rho(s)$.

If we introduce the *spectral density function*

$$F(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} e^{i\boldsymbol{\omega} \cdot \mathbf{x}} \rho(\mathbf{x})\, d\mathbf{x} \tag{8}$$

and the *areal characteristic function*

$$G(\boldsymbol{\omega}) = \int_\Omega e^{i\boldsymbol{\omega} \cdot \mathbf{x}} d\mathbf{x}, \tag{9}$$

then equation (4) can be rewritten as

$$V(\Omega) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} |\, G(\boldsymbol{\omega})\, |^2 F(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \tag{10}$$

This formula sometimes has advantages over (4), since the limits of integration are simpler, the weights given to different wave numbers $\boldsymbol{\omega}$ are quite evident, and approximate methods

such as the method of steepest descents can be used immediately. The function $G(\boldsymbol{\omega})$ is readily evaluated for simple plot shapes. Thus for the rectangle with sides a, b

$$G(\boldsymbol{\omega}) = \frac{4\sin(a\omega_1)\sin(b\omega_2)}{\omega_1\omega_2}, \tag{11}$$

and for a circle with radius a and centre at the origin

$$G(\boldsymbol{\omega}) = \frac{2\pi a}{\omega} J_1(a\omega), \tag{12}$$

where J_1 is the Bessel function of order 1. Here ω_1 and ω_2 are the components of $\boldsymbol{\omega}$, and ω is its absolute value.

3. THE INVERSE PROBLEM

From the practical point of view, it would be useful to be able to invert relation (4), and express the covariance function $\rho(\mathbf{s})$ in terms of the yield variance $V(\Omega)$, which can presumably be observed for varying Ω.

This inversion is very simple for the special cases (5) and (6), corresponding to linear and rectangular plots. We have respectively

$$\rho(x) = \frac{1}{2}\frac{\partial^2 V(x)}{\partial x^2}, \tag{13}$$

$$\rho(x, y) = \frac{1}{4}\frac{\partial^4 V(x, y)}{\partial x^2 \partial y^2}. \tag{14}$$

However, there are few cases as tractable as these.

We shall now largely confine ourselves to one general type of case. We shall assume that the covariance is *isotropic*, so that $\rho(\mathbf{s})$ is a function of s only, and we may write

$$\rho(\mathbf{s}) = \rho(s). \tag{15}$$

Further, we shall assume that we are dealing with a plot of constant shape, which can only vary similarly, i.e. by changing all its dimensions in the same ratio. For a given plot shape, the size of the plot is conveniently specified by its largest dimension, x (say), and we shall write

$$V(\Omega) = V(x). \tag{16}$$

(Thus, for a circular plot x equals the diameter, for a rectangular plot x equals the diagonal, etc.) Now, corresponding to each plot shape there will be a function $K(s)$ describing the distribution of distances between two points chosen randomly in the plot, such that

$$V(1) = \int_0^1 K(s)\rho(s)\,ds. \tag{17}$$

If we now increase all linear dimensions in the ratio $x:1$, then the infinitesimal elements of integration will be increased in the ratio $x^n:1$, and we shall have quite generally

$$V(x) = x^{2n}\int_0^1 K(s)\rho(xs)\,ds. \tag{18}$$

The form of the relation (18) between V and ρ invites us to introduce the Mellin transforms

$$\overline{V}(\nu) = \int_0^\infty V(s)\, s^{\nu-1} ds, \tag{19}$$

$$\overline{\rho}(\nu) = \int_0^\infty \rho(s)\, s^{\nu-1} ds, \tag{20}$$

$$\overline{K}(\nu) = \int_0^1 K(s)\, s^{\nu-1} ds, \tag{21}$$

in terms of which (18) becomes

$$\overline{V}(\nu-2n) = \overline{K}(1-\nu)\,\overline{\rho}(\nu). \tag{22}$$

We shall consider in § 4 the common range of ν for which the transforms (19) and (20) exist, and (22) is valid. Assuming for the moment that a suitable range exists, we can immediately obtain a solution for $\rho(s)$ by taking the inverse transform of the expression for $\overline{\rho}$ provided by (22). In fact, if the desired inverse relation to (18) is

$$\rho(x) = x^{-2n} \int_0^1 L(s)\, V(xs)\, ds, \tag{23}$$

then

$$\overline{\rho}(\nu) = \overline{L}(1+2n-\nu)\,\overline{V}(\nu-2n), \tag{24}$$

where $\overline{L}(\nu)$ has a similar definition to $\overline{K}(\nu)$. Comparing (22) and (24) we see that

$$\overline{L}(\nu) = \frac{1}{\overline{K}(\nu-2n)}. \tag{25}$$

In order to be able to carry out the calculation explicitly we must then be able to perform the integration (21) and then invert the $\overline{L}(\nu)$ yielded by (25). One of the few cases for which this is possible is the rather unrealistic one of a circular plot (x = diameter, $n=2$). For the circle we obtain by direct methods

$$K(s) = \tfrac{1}{2}[s\cos^{-1}(s) - s^2\sqrt{(1-s^2)}], \tag{26}$$

whence

$$\overline{K}(\nu) = \frac{\Gamma(\frac{1}{2})\,\Gamma\{\frac{1}{2}(\nu+2)\}}{4(1+\nu)\,\Gamma\{\frac{1}{2}(\nu+5)\}}, \tag{27}$$

$$\overline{L}(\nu) = \frac{4(\nu-3)\,\Gamma\{\frac{1}{2}(\nu+1)\}}{\Gamma(\frac{1}{2})\,\Gamma\{\frac{1}{2}(\nu-2)\}}. \tag{28}$$

However, upon attempting to invert $\overline{L}(\nu)$, we obtain a divergent integral. This may be taken as an indication of the fact that $\rho(x)$ should be expressed not only in terms of $V(x)$ but also of its derivatives $V^{(j)}(x)$. In view of results (13) and (14) this is not surprising. If for (23) we substitute the more general relation

$$\rho(x) = x^{-2n} \int_0^1 \sum_{j=0}^{\infty} L_j(s)\,[(xs)^j\, V^{(j)}(xs)]\, ds, \tag{29}$$

we find that, provided that

$$[s^{\nu+j-1} V^{(j)}(s)]_0^\infty = 0 \quad (j = 0, 1, 2, \ldots), \tag{30}$$

then

$$\overline{L}(\nu) = \sum_j (\nu-1)(\nu-2)\ldots(\nu-j)\,\overline{L}_j(\nu). \tag{31}$$

Now, the $\overline{L}(\nu)$ of (28) can be written

$$\pi^{-1} B\left(\frac{1}{2}, \frac{\nu-3}{2}\right)[(\nu-1) - (\nu-1)(\nu-2) + (\nu-1)(\nu-2)(\nu-3)], \tag{32}$$

which, in view of (29) and the integral representation of the Beta-function, corresponds to the relation

$$\rho(x) = \frac{2x^{-4}}{\pi}\int_0^1 [(xs)\,V'(s) - (xs)^2\,V''(xs) + (xs)^3\,V'''(xs)]\frac{s^{-3}ds}{\sqrt{(1-s^2)}}. \tag{33}$$

A case of greater practical interest would be that of the rectangle. However, while $K(s)$ is easily calculated for this figure, the transform $\bar{K}(\nu)$ is not.

4. Power-law covariances

Let us first return to relation (22). If a function $f(s)$ is $O(s^{-\alpha})$ and $O(s^{-\beta})$ for small and large positive s respectively, then its Mellin transform $\bar{f}(\nu)$ will certainly exist for

$$\alpha < \mathrm{Re}\,(\nu) < \beta \tag{34}$$

and $\bar{f}(\nu)$ will in fact have simple poles at α and β.

Now, $\rho(s)$ is $O(1)$ at the origin (if we suppose it continuous there), and if we assume it is $O(s^{-\lambda})$ for large ν, then $\bar{\rho}(\nu)$ exists for $0 < \mathrm{Re}\,(\nu) < \lambda$. For small s, $K(s)$ is proportional to the surface area of an n-dimensional sphere of radius s, and is hence $O(s^{n-1})$, while for $s > 1$ it is zero. Consequently $\bar{K}(\nu)$ exists for $(1-n) < \mathrm{Re}\,(\nu) < \infty$.

It thus follows that both quantities in the right-hand member of (22) exist for

$$0 < \mathrm{Re}\,(\nu) < \min\,(n, \lambda). \tag{35}$$

Since relation (22) determines $\bar{V}$ in terms of $\bar{K}$ and $\bar{\rho}$, the relation is also valid for these values of ν, and $\bar{V}(\nu)$ exists for

$$-2n < \mathrm{Re}\,(\nu) < \min\,(-n, \lambda - 2n), \tag{36}$$

and has simple poles at these extreme values of ν.

We can thus conclude that for small plots

$$V(x) = O(x^{2n}) = O(A^2), \tag{37}$$

while for large plots

$$V(x) = O(x^n) = O(A), \tag{38}$$

if $\rho(s)$ falls off more rapidly than s^{-n}, or

$$V(x) = O(x^{2n-\lambda}) = O(A^{2-\lambda/n}), \tag{39}$$

if $\rho(s)$ falls off as $s^{-\lambda}$ $(\lambda < n)$. In other words, it is a power law $[\rho(s) \sim Cs^{-n}]$ which marks the transition between the two cases where V is proportional to plot area for large plots, or increases faster than plot area.

Relation (22) may be used in reverse: if $V(x)$ is observed to increase as A^μ for large areas $(1 < \mu < 2)$, then we can deduce that $\rho(s)$ falls off as $s^{n(\mu-2)}$ for large distances.

It should be emphasized that these conditions hold only for plots of constant shape, since V will in general be a function of shape as well as area. However, some results presented by P. Fairfield Smith (1938) would seem to indicate that this dependence on shape may be weak, and that V may be determined almost entirely by plot area. This is presumably true only if the shape is not too extreme, e.g. not too narrow and elongated.

Fairfield Smith's results are also of interest in that they provide very convincing evidence of the behaviour associated with equation (39). He found that to a very good approximation

the variance of yield per unit area in plots of area A could be represented as a curve const. $A^{-0.749}$. Replacing the index by $-\frac{3}{4}$, we then have

$$V(A) \sim \text{const.}\, A^{2-\frac{3}{4}} = \text{const.}\, A^{\frac{5}{4}}, \tag{40}$$

corresponding to a covariance function

$$\rho(s) \sim \text{const.}\, s^{-\frac{3}{2}}. \tag{41}$$

(These equalities cannot, of course, hold for indefinitely small A and s.)

This result, well founded in observation, provides evidence of three intriguing possibilities: (*a*) that covariances decaying as $s^{-\lambda}$ do occur in nature; (*b*) that the rate of decay may be so small that yield variance increases faster than plot area; (*c*) that the observed index λ may have simple rational values.

None of the simple linear models hitherto proposed to represent yield variation over an area (see Whittle, 1954; Heine, 1955) lead to covariances of the above type. For instance, the model which relates yield density $\xi(x, y)$ symmetrically to the yield at all points around (x, y):

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - \kappa^2\right] \xi(x, y) = \epsilon(x, y) \tag{42}$$

leads to the covariance function

$$\begin{aligned} \rho(s) &= \text{const.}\, sK_1(\kappa s) \\ &\sim \text{const.}\, s^{\frac{1}{2}} e^{-\kappa s} \quad (\kappa s \text{ large}). \end{aligned} \tag{43}$$

(Here K_1 denotes a modified Bessel function.)

For the one-dimensional problem of yarn diameter variation D. R. Cox (personal communication regarding unpublished work) has proposed a model which does in fact predict a power-law covariance for large s. The only multi-dimensional models known to the author which predict a power-law covariance for any range of s are those of turbulence (see Batchelor, 1953, p. 122). For these models dimensional arguments show that the index λ must have simple rational values.

It seems likely to the author that any model which is to provide a satisfactory explanation of the power-law decay observed in agricultural work must embody two of the features common to Cox's yarn model and the turbulence model: (*a*) it must be non-linear, and (*b*) it must consider the variate (yield density) to be a function of time as well as of the spatial co-ordinates. (In Cox's case 'time' is discrete, the number of smoothings that the yarn has undergone.) As an example of such a model, one could suppose that fertility gradients in the soil tend to be smoothed out in the course of time by a diffusion process, which is non-linear to the extent that only gradients which are greater than a certain value tend to diminish.

The value of the index λ deserves some discussion. Is it fortuitous that Fairfield Smith's estimated index $b = 0{\cdot}749$ lies so near the rational value $\frac{3}{4}$ (corresponding to $\lambda = \frac{3}{2}$)? Certainly little can be inferred from this single instance. However, Fairfield Smith lists the b values estimated from uniformity data for several different kinds of crop. The histogram of the thirty-nine estimated b values has its main peak in the interval $b = 0{\cdot}41$–$0{\cdot}50$, and a second peak and upper cut-off point at $b = 0{\cdot}71$–$0{\cdot}80$. This provides at least a suggestion that the

values $b = \frac{1}{2}$ and $\frac{3}{4}$ (corresponding to $\lambda = 1, \frac{3}{2}$) are in some way distinguished. Perhaps one should say no more until further data or predictions from theory are available.

The author is grateful to Dr D. R. Cox for drawing his attention to the observations on yarn diameter and flood height.

REFERENCES

BATCHELOR, G. K. (1953). *The Theory of Homogeneous Turbulence*. Cambridge University Press.
FAIRFIELD SMITH, H. (1938). *J. Agric. Sci.* **28**, 1–23.
FELLER, W. (1951). The asymptotic distribution of the range of independent random variables. *Ann. Math. Statist.* **22**, 427–32.
HEINE, V. (1955). Models for two-dimensional stationary stochastic processes. *Biometrika*, **42**, 170–8.
WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434–49.

ON THE CONSTRUCTION OF SIGNIFICANCE TESTS ON THE CIRCLE AND THE SPHERE

By G. S. WATSON* and E. J. WILLIAMS†

1. Introduction

A number of recent papers have dealt with the probability density, in two and three dimensions, proportional to

$$\exp(\kappa \cos\theta),$$

where κ is a precision constant, and θ is the angle between an observed unit vector and the population mean unit vector or polar vector. The purposes of these investigations have been (i) to derive, from observed results, limits within which the unknown polar vector is likely to lie, and (ii) to test the homogeneity of different sets of observations, both in their precision and in their direction. These distributions and the tests associated with them thus produce the circular and spherical analogues of the usual tests in Euclidean space, based on the normal distribution.

For example, in palaeo-magnetism, a sample of rock specimens is collected from a site and the direction of remanent magnetism of each specimen is measured. The directions may be loosely or closely grouped about a mean direction. It has been shown (Irving & Watson, 1956) that the above distribution is a good fit in the case of stably magnetized rocks, for both large and small dispersions. From the knowledge of the mean direction of magnetization, the geophysicist hopes to learn about the earth's magnetic field at the time the rocks were magnetized. Such knowledge is being used to throw light on the stability or otherwise of the earth's magnetic axis and on the positions of the continents in different geological epochs.

Fisher (1953) considered the three-dimensional case where the observations may be regarded as points on a sphere. He derived the maximum-likelihood estimates of κ and the polar direction and provided the basic distribution theory which he used to find the fiducial test of a prescribed polar direction when κ is unknown. Watson (1956*a*, *b*) gave a significance table for the test of $\kappa = 0$, and approximate tests for the equality of the κ's and polar directions for any number of populations.

Gumbel, Greenwood & Durand (1953) and Greenwood & Durand (1955) studied the two-dimensional case where the observations may be regarded as points on the unit circle. They gave a table to facilitate the calculation of the maximum-likelihood estimate of κ and a significance table for testing $\kappa = 0$. For this case, less is known of the distribution theory. Pearson (1906), Kluyver (1906) and Rayleigh (1919) gave results on the distribution of the length of the resultant vector of a sample when $\kappa = 0$. Greenwood & Durand gave some further results when $\kappa \neq 0$.

In the present paper, the fundamental property of sufficient statistics will be used to derive tests, free of nuisance parameters, for direction and homogeneity in both the two- and three-dimensional cases. In only one situation, however, can we give the required exact distribution. Inequalities are suggested for some of the tests proposed; and an arithmetical example suggests that these may be useful in practical applications.

* The Australian National University, Canberra, A.C.T. † C.S.I.R.O., Melbourne.

2. BASIC RESULTS

We write the density function as

$$C_p(\kappa)\, e^{\kappa \cos\theta}, \tag{1}$$

with p the number of dimensions and $C_p(\kappa)$ the corresponding constant factor. In general, the range of θ is

$$0 \leqslant \theta < \pi,$$

but in the two-dimensional case it is conventional (though not necessary) to take

$$0 \leqslant \theta < 2\pi.$$

The constant factor is, in general, the reciprocal of

$$2\frac{\pi^{\frac{1}{2}(p-1)}}{\Gamma(\frac{1}{2}(p-1))}\int_0^\pi \sin^{p-2}\theta\, e^{\kappa\cos\theta}\, d\theta, \tag{2}$$

which may be expressed in terms of the imaginary Bessel function $I_{\frac{1}{2}p-1}(\kappa)$ or (when p is odd) in terms of $\sinh\kappa$ and $\cosh\kappa$. When $p=2$ the factor must be halved. Then

$$C_2(\kappa) = 1/(2\pi I_0(\kappa)) \tag{3}$$

and

$$C_3(\kappa) = \kappa/(4\pi\sinh\kappa). \tag{4}$$

The probability density of a sample of N is a function only of

$$\sum_{i=1}^{N}\cos\theta_i = X.$$

Hence, when the polar vector is assumed known, so that the θ_i can be determined from the observed vectors, X is a sufficient statistic for κ, the maximum-likelihood (ML) equations for κ being

$$\frac{I_0'(\kappa)}{I_0(\kappa)} = \frac{X}{N} \quad (p=2) \tag{5}$$

and

$$\coth\kappa - \frac{1}{\kappa} = \frac{X}{N} \quad (p=3). \tag{6}$$

When the polar vector is not known, its ML estimator is, for all values of p, the set of direction cosines of the vector resultant of the sample of N unit vectors; the ML equations for κ are (5) and (6) (in the cases $p=2$ or 3) with X replaced by R, the length of the resultant.

If c is the cosine of the angle between the resultant and polar vectors, then $X = Rc$; R is determinable but not c unless the polar vector is known. From the form of the distribution it is seen that Rc is a sufficient statistic for κ. Hence the joint probability density of R and c, when $\kappa \neq 0$, is found by multiplying the joint probability density, when $\kappa = 0$, by the factor

$$[C_p(\kappa)]^N e^{\kappa Rc}.$$

We therefore first consider the uniform distribution given by $\kappa = 0$. When $\kappa = 0$, R and c are independently distributed.

In two dimensions, the density of R may be expressed as an integral (Kluyver, 1906)

$$R\int_0^\infty [J_0(t)]^N J_0(Rt)\, t\, dt, \tag{7}$$

or as an asymptotic series (Pearson, 1906). These representations are discussed by Greenwood & Durand (1955). Hence, in this case, the joint density of R and c is given by

$$\frac{e^{\kappa Rc}}{[I_0(\kappa)]^N} R\left\{\int_0^\infty [J_0(t)]^N J_0(Rt)\, t\, dt\right\} dR\, \frac{dc}{\pi(1-c^2)^{\frac{1}{2}}}, \tag{8}$$

which is the same as formula (3·6) of Greenwood & Durand.

In three dimensions, Fisher gives the joint density

$$\left(\frac{\kappa}{2\sinh\kappa}\right)^N e^{\kappa Rc}\,\phi_N(R)\,R\,dR\,dc, \tag{9}$$

where $$\phi_N(R) = \frac{1}{(N-2)!}\sum_{s=0}^{\infty}\binom{N}{s}(-1)^s\langle N-R-2s\rangle^{N-2},$$

with the notation $\langle x\rangle = x$ if $x \geqslant 0$, $\langle x\rangle = 0$ if $x < 0$.

It follows from these results, or can be derived directly, that the density of $X = Rc$ in two dimensions is

$$\frac{e^{\kappa X}}{\pi[I_0(\kappa)]^N}\int_0^{\infty}[J_0(t)]^N\cos Xt\,dt, \tag{10}$$

and in three dimensions is

$$\left(\frac{\kappa}{2\sinh\kappa}\right)^N\frac{e^{\kappa X}}{(N-1)!}\sum_{s=0}^{\infty}\binom{N}{s}(-1)^s\langle N-X-2s\rangle^{N-1}. \tag{11}$$

For confidence limits on κ, when the polar directions are known, and for tests of a prescribed polar vector, the probability densities (10) and (11) suffice, since in either case the null hypothesis specifies c so that X is determinable.

To test the homogeneity of the values of κ for several sets of results, or to compare polar vectors, the same value of κ being assumed, we need to work with the sample resultants. In such cases the joint distribution of the sample resultants and their overall resultant is required. We suppose that there are samples of sizes $N_1, N_2, \ldots$ with resultants of length $R_1, R_2, \ldots$ and that the overall resultant has length R.

For resultants in two dimensions, the joint distribution is not easy to determine. We give here the results only for two samples. If the integral in (8) is denoted by $\psi_N(R)$, then the joint density of R_1, c_1, R_2 and c_2 from two independent samples is

$$\frac{\exp\{\kappa(R_1c_1+R_2c_2)\}}{[I_0(\kappa)]^{N_1+N_2}}\,\frac{R_1\psi_{N_1}(R_1)\,R_2\psi_{N_2}(R_2)}{\pi^2\{(1-c_1^2)\,(1-c_2^2)\}^{\frac{1}{2}}}. \tag{12}$$

We put $R_1c_1 + R_2c_2 = Rc$. Then the joint distribution of R_1, R_2, R and c may be found by the following simple geometrical argument.

Since Rc is sufficient for κ, the factor introducing κ may be ignored in the derivation, and brought back into the final result. Thus κ may be assumed zero, so that the angles are uniformly distributed independently of the R's.

R^2 ranges from $(R_1-R_2)^2$ to $(R_1+R_2)^2$, being given by

$$R^2 = (R_1-R_2)^2\cos^2\lambda + (R_1+R_2)^2\sin^2\lambda,$$

where λ is half the angle between the two vectors. Also, since λ is uniformly distributed from 0 to π, we may determine the conditional distribution of R given R_1, R_2.

For $$RdR = 2R_1R_2\sin 2\lambda\,d\lambda,$$

while $$\sin 2\lambda = \{[(R_1+R_2)^2-R^2]\,[R^2-(R_1-R_2)^2]\}^{\frac{1}{2}}/(2R_1R_2),$$

so that $$RdR = \{[(R_1+R_2)^2-R^2]\,[R^2-(R_1-R_2)^2]\}^{\frac{1}{2}}\,d\lambda.$$

Hence the elementary conditional probability is

$$\frac{d\lambda}{\pi} = \frac{RdR}{\pi\{[(R_1+R_2)^2-R^2]\,[R^2-(R_1-R_2)^2]\}^{\frac{1}{2}}}$$
$$= G(R_1, R_2, R)\,dR \quad \text{(say)}.$$

The joint probability density of R_1, R_2, R and c (still with $\kappa = 0$) is therefore

$$\frac{R_1 \psi_{N_1}(R_1) R_2 \psi_{N_2}(R_2) G(R_1, R_2, R)}{\pi(1-c^2)^{\frac{1}{2}}}. \tag{13}$$

Bringing in the factor introducing κ, we find the non-null joint density to be $(N = \Sigma N_i)$

$$\frac{e^{\kappa R c}}{[I_0(\kappa)]^N} \frac{R_1 \psi_{N_1}(R_1) R_2 \psi_{N_2}(R_2) G(R_1, R_2, R)}{\pi(1-c^2)^{\frac{1}{2}}}. \tag{14}$$

Finally, integrating out c, we have for the joint density of R_1, R_2 and R

$$\frac{I_0(\kappa R)}{[I_0(\kappa)]^N} R_1 \psi_{N_1}(R_1) R_2 \psi_{N_2}(R_2) G(R_1, R_2, R). \tag{15}$$

The density of R is found, by integrating (8) over all values of c, as

$$\frac{I_0(\kappa R)}{[I_0(\kappa)]^N} R \psi_N(R). \tag{16}$$

Hence the conditional density of R_1 and R_2, given R, is

$$\frac{R_1 \psi_{N_1}(R_1) R_2 \psi_{N_2}(R_2) G(R_1, R_2, R)}{R \psi_N(R)}. \tag{17}$$

Being independent of κ, this conditional density provides the basis for exact significance tests. So far it has not been possible to generalize this result to more than two samples.

For resultants in three dimensions, Fisher (1953) has shown that, for any number of samples, the joint-probability density of the R_i and R is

$$\left(\frac{\kappa}{2\sinh\kappa}\right)^N \frac{2\sinh\kappa R}{\kappa} \Pi \phi_{N_i}(R_i), \tag{18}$$

while the density of R is

$$\left(\frac{\kappa}{2\sinh\kappa}\right)^N \frac{2\sinh\kappa R}{\kappa} \phi_N(R), \tag{19}$$

so that the conditional density of the R_i, given R, is

$$\Pi \phi_{N_i}(R_i) / \phi_N(R). \tag{20}$$

This conditional density is the basis for any exact significance tests.

3. TESTS OF SIGNIFICANCE

The required significance tests for data following circular or spherical distributions have been discussed in § 1.

When κ is large we may make use of the fact that $2\kappa(1-\cos\theta)$ is distributed approximately as χ^2 with $p-1$ degrees of freedom. Hence, we have approximately

$$\left.\begin{aligned} 2\kappa(N-X) &= \chi^2_{(p-1)N}, \\ 2\kappa(N-R) &= \chi p^2_{(p-1)(N-1)}. \end{aligned}\right\} \tag{21}$$

We may therefore take $N - \Sigma R_i$ and $\Sigma R_i - R$ to be independently distributed as χ^2 with $(p-1)(N-q)$ and $(p-1)(q-1)$ degrees of freedom respectively, where q is the number of samples.

The tests based on these results were discussed for the spherical case by Watson (1956*a*). They will certainly be accurate when κ is large, but seem to be accurate also for a wide range of κ and N. As is usual in problems of this kind, it is hard to assess the degree of approximation of (21). In § 4 some more results will be given to justify the tests derived from (21).

To test a given value of κ or to derive confidence limits for κ, we should use

$$2\kappa(N-X), \quad \text{when the polar vector is known, or}$$
$$2\kappa(N-R), \quad \text{when the polar vector is unknown.}$$

Likewise, to test the homogeneity of κ-values for different samples, we should compare corresponding values of $N-X$ or $N-R$.

To test the equality of polar vectors for different samples, we should use, generalizing Watson (1956*a*),

$$\frac{(N-q)(\Sigma R_i - R)}{(q-1)(N-\Sigma R_i)}, \tag{22}$$

which is distributed as F with $(p-1)(q-1)$ and $(p-1)(N-q)$ degrees of freedom.

The distributions of all the test statistics given above carry κ as a nuisance parameter. The tests given below do not, but the required distributions, with one exception, are not known. However, some inequalities are given which may help in their application.

(*a*) *Tests of hypotheses concerning κ's*

It has been seen in § 2 that when the polar vector is known, $X = \Sigma \cos\theta$ is a sufficient statistic for κ. To test that several populations with known polar vectors have a common value of κ, suppose samples of $N_1, N_2, \ldots$ with values $X_1, X_2, \ldots$ are available. Then the joint density of $X_1, X_2, \ldots$, given the value X of ΣX_i, may be written as

$$\frac{\lambda_{N_1}(X_1)\lambda_{N_2}(X_2)\ldots}{\lambda_N(X)}, \tag{23}$$

where $\lambda_N(X)$ stands for the factor of (10) or (11) independent of κ. The distribution of any test function found from the density (23) will be independent of κ. However, as the case of a known polar vector is of no practical interest, it will not be pursued further.

When the polar vectors are unknown, the test of the equality of κ in several populations must be based on the resultants $R_1, R_2, \ldots$ of the several samples, and the overall resultant R. Formulae (17) or (20) show that the joint density of $R_1, R_2, \ldots$ given R is independent of κ, so that an exact test can be made. When κ is large, or when the N_i are all large, the test function must be sensitive to deviations of the ratios R_i/N_i from equality. Under these conditions, the tests based on (21) are likely to be adequate. Nothing is known of suitable tests when both κ and the N_i are small.

(*b*) *Tests of hypotheses concerning polar vectors*

(i) *A single prescribed polar vector.* If κ is known, a test of a prescribed polar vector may be made most conveniently by using the density of X ((10) or (11)). In practice κ is not known. To derive an analogue to the single-sample Student t-test, we consider the density of R given X. Since the density of X is an even function of X, these are strictly tests of a prescribed axis. In practice, however, the observations will usually be so grouped that no confusion will arise.

In two dimensions, this is found from (7) and (10) to be

$$\frac{R\int_0^\infty [J_0(t)]^N J_0(Rt)\,t\,dt}{\int_0^\infty [J_0(t)]^N \cos Xt\,dt\,(R^2-X^2)^{\frac{1}{2}}}, \tag{24}$$

where $R \geqslant X$. If the sample mean direction is far from the prescribed direction, R will be much greater than X. Thus to define a possible significance test, we find a value of R, R_0 say, so that

$$P(R > R_0 \,|\, X) = \alpha,$$

i.e. so that

$$\int_{R_0}^N \frac{R}{(R^2-X^2)^{\frac{1}{2}}}\,dR \int_0^\infty [J_0(t)]^N J_0(Rt)\,t\,dt = \alpha \int_0^\infty [J_0(t)]^N \cos Xt\,dt. \tag{25}$$

This is difficult to solve for R_0. Since, however, the integral of (24) from X to N is unity, we have the relation

$$\begin{aligned}\int_0^\infty [J_0(t)]^N \cos R_0 t\,dt &= \int_{R_0}^N \frac{R}{(R^2-R_0^2)^{\frac{1}{2}}}\,dR \int_0^\infty [J_0(t)]^N J_0(Rt)\,t\,dt \\ &\geqslant \int_{R_0}^N \frac{R}{(R^2-X^2)^{\frac{1}{2}}}\,dR \int_0^\infty [J_0(t)]^N J_0(Rt)\,t\,dt,\end{aligned}$$

since $R \geqslant R_0 \geqslant X$. Thus, in two dimensions,

$$P(R > R_0 \,|\, X) \leqslant \frac{\int_0^\infty [J_0(t)]^N \cos R_0 t\,dt}{\int_0^\infty [J_0(t)]^N \cos Xt\,dt}. \tag{26}$$

The functions on the right-hand side of (26) could, with some difficulty, be evaluated numerically for various R, X and N or, more approximately, be replaced by their saddle-point approximations or by their approximations when R and X are large, i.e. near N.

In three dimensions, we have instead of (24)

$$\frac{\phi_N(R)}{\frac{1}{(N-1)!}\sum_{s=0}^\infty \binom{N}{s}(-1)^s \langle N-X-2s\rangle^{N-1}}, \tag{27}$$

so that, by an easy calculation,

$$P(R > R_0 \,|\, X) = \frac{\sum_{s=0}^\infty \binom{N}{s}(-1)^s \langle N-R_0-2s\rangle^{N-1}}{\sum_{s=0}^\infty \binom{N}{s}(-1)^s \langle N-X-2s\rangle^{N-1}}. \tag{28}$$

It will be seen that the right-hand sides of (26) and (28) have the same form, i.e. the ratio of the density of X evaluated at R and at X. Furthermore, the ratio of the leading terms in (28),

$$\left(\frac{N-R}{N-X}\right)^{N-1} = \left(\frac{N-R}{N-Rc}\right)^{N-1},$$

is Fisher's first approximation to the probability of a cosine less than c, calculated fiducially. In computing (28), this factor is best taken separately in order to reduce the number of working figures.

(ii) *Comparison of several polar vectors.* It will often be required to test the equality of the polar vectors of several populations. To devise a possible test statistic, we observe that, if the sample resultants have very similar orientations, $R_1+R_2+\ldots$ will not be very much greater than R, while if the orientations are not alike $R_1+R_2+\ldots$ will be much larger than R. Since the density of $R_1+R_2+\ldots$ given R is free of κ, it provides a possible exact test.

In the case of two populations in two dimensions, the joint density of R_1 and R_2, given R, is given by (17), where the region of joint variation of R_1, R_2 and R is defined by

$$\left.\begin{aligned} 0 \leqslant R_1 \leqslant N_1, \quad 0 \leqslant R_2 \leqslant N_2, \quad 0 \leqslant R \leqslant N, \\ |R_1 - R_2| \leqslant R \leqslant R_1 + R_2. \end{aligned}\right\} \tag{29}$$

We require a value R_0 such that

$$P(R_1+R_2 > R_0 \,|\, R) = \alpha. \tag{30}$$

As with equation (25), it seems difficult to solve for R_0, but an inequality is easily obtained by the method used to derive (26). Thus

$$P(R_1+R_2 > R_0 \,|\, R) \leqslant \frac{R_0 \psi_N(R_0)}{R \psi_N(R)}. \tag{31}$$

For application of (31), the function ψ_N must be found. Use may be made of the fact that $\psi_N(R)$ is asymptotically $2\exp(-R^2/N)/N$.

An alternative test is given by the function

$$N(N-1)(N-2)\frac{[(R_1+R_2)^2-R^2][R^2-(R_1-R_2)^2]}{(N^2-R^2)[N^2R^2-(R_1^2-R_2^2)^2]}, \tag{32}$$

which is distributed approximately as χ^2 with 1 degree of freedom. This tests equality or oppositeness of direction; that is, departures of R either from R_1+R_2 or from $|R_1-R_2|$.

For several samples in three dimensions, the joint density of $R_1, R_2, \ldots$ given R is given by (20). Once again, only an inequality is easily found. This is that

$$P(\Sigma R_i > R_0 \,|\, R) \leqslant \frac{\phi_N(R_0)}{\phi_N(R)}. \tag{33}$$

Here the function $\phi_N(R)$ is easy to compute in any application. Actually there are cases where (33) is an equality. For example, with two samples only, the domain of R_1 and R_2 given R is (29). A little consideration shows that there will be equality in (33) if $R \geqslant \max(N_i)$ or if $R_0+R \geqslant \max(2N_i)$.

4. Numerical application and comparison of significance tests

In this section, we apply some of the tests suggested here to the numerical examples given by Fisher (1953) and Watson (1956*a*) and compare the results with those obtained in these papers. The agreement is remarkable, and, since Watson's tests are the easiest to apply, it is suggested that they may be used with confidence despite their approximate derivation.

To begin with we consider Fisher's example (*a*). It is required to find the 5% zone of confidence for the polar direction. A sample of $N=9$ is available and $R=8{\cdot}77203$. Fisher uses his approximate form for the probability that the cosine of the angle between the resultant and the polar vector is less than c,

$$P = \left(\frac{N-R}{N-Rc}\right)^{N-1}, \tag{34}$$

with $P = 0{\cdot}05$. He finds $c = 0{\cdot}98820$ so that the confidence zone includes only directions less than 8·8° away from the resultant direction. We have shown that Fisher's solution is, in general, a first approximation to our exact expression (28). Here, in fact, they coincide as R is so close to N. So Fisher's solution is also that obtained from (28). Watson's approximate result in this case is that

$$(N-1)\frac{R(1-c)}{N-R} \approx F_{2,2(N-1)}, \tag{35}$$

which is equivalent to Fisher's result (34). Incidentally, the estimate of κ from these data is 35·1.

In Fisher's example (*b*), a sample of 45 more widely dispersed ($\hat{\kappa} = 7{\cdot}3$) directions is examined. Since $R = 38{\cdot}9946$, Fisher finds that the 5 % zone of confidence is limited by $c = 0{\cdot}98915$, i.e. $\theta = 8{\cdot}45°$, by his formula given above. Taking $F_{2,88}(5\,\%) = 3{\cdot}10$, formula (35) gives an identical value of c. The extra terms in (28) and also the extra terms in Fisher's more complicated formula for P make no difference to the value of c. Thus the approximation (34) is entirely satisfactory for small values of κ and medium values of N.

Needless to say, had the various formulae been used for a test of significance of a given polar direction on the above data, similar agreements would have been found.

Watson gives an example of a test that the polar directions of three populations are identical. In the data analysed

	Sample 1	Sample 2	Sample 3
N_i	10	11	15
R_i	6·990	8·212	12·194

while for the combined sample $N = 36$, $R = 26{\cdot}902$. In this situation, Watson proposed the approximation

$$F_{4,2(N-3)} = \frac{(N-3)}{2}\frac{\Sigma R_i - R}{N - \Sigma R_i}. \tag{36}$$

The right-hand side is found to be 1·81. Graphical interpolation in the F-tables show that $F = 1{\cdot}81$ corresponds to a probability of about 13 %. The inequalty (33) of this paper may be applied to this problem by taking

$$R_0 = \Sigma R_i = 27{\cdot}396, \quad R = 26{\cdot}902.$$

Now

$$\frac{\phi_N(R_0)}{\phi_N(R)} = \left(\frac{N-R_0}{N-R}\right)^{N-2} \frac{\sum_{s=0}^{\infty} (-1)^s \binom{N}{s} \left\langle \frac{N-R_0-2s}{N-R_0} \right\rangle^{N-2}}{\sum_{s=0}^{\infty} (-1)^s \binom{N}{s} \left\langle \frac{N-R-2s}{N-R} \right\rangle^{N-2}}$$

$$= 0{\cdot}1499 \times 1{\cdot}0034$$

$$= 0{\cdot}1504.$$

Since 0·15 is only slightly greater than 0·13, it seems likely that very little is lost in forming the inequality (33) (i.e. it is close to equality) in practice and that the F-approximation is again accurate.

Finally, we notice in this case that the values of N and $\hat{\kappa}$ in each sample are both small. Furthermore, the value of $\phi_N(R_0)/\phi_N(R)$ is accurately given by the leading term so that one might suggest the simple approximation

$$P(\Sigma R_i > R_0 \,|\, R) \approx \left(\frac{N-R_0}{N-R}\right)^{N-2}. \tag{37}$$

Assuming that the approximations (21) are adequate for testing purposes, the relation (37) has the circular analogues

$$P(R > R_0 | X) \approx \left(\frac{N - R_0}{N - X}\right)^{\frac{1}{2}N-1} \tag{38}$$

and

$$P(\Sigma R_i > R_0 | R) \approx \left(\frac{N - R_0}{N - R}\right)^{\frac{1}{2}(N-1)-1}. \tag{39}$$

The merits of these approximations are unknown.

REFERENCES

Fisher, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc.* A, **217**, 295.

Greenwood, J. A. & Durand, D. (1955). The distribution of length and components of the sum of random unit vectors. *Ann. Math. Statist.* **26**, 233.

Gumbel, E. J., Greenwood, J. A. & Durand, D. (1953). The circular normal distribution: theory and tables. *J. Amer. Statist. Ass.* **48**, 131.

Irving, E. & Watson, G. S. (1956). The use of statistics in studies of the magnetic directions in rocks (unpublished).

Kluyver, J. C. (1906). A local probability problem. *Proc. Acad. Sci. Amst.* **8**, 341.

Pearson, K. (1906). A mathematical theory of random migration. *Drap. Co. Res. Mem.* no. 3.

Strutt, J. W. (Lord Rayleigh) (1919). On the problem of random vibrations and random flights in one, two and three dimensions. *Phil. Mag.* (6), **37**, 321.

Watson, G. S. (1956*a*). Analysis of dispersion on a sphere. *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **7**, 153.

Watson, G. S. (1956*b*). A test for randomness of directions. *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **7**, 160.

NOTES ON BIAS IN ESTIMATION

By M. H. QUENOUILLE

Research Techniques Unit, London School of Economics and Political Science

1. One of the commonest problems in statistics is, given a series of observations $x_1, x_2, \ldots, x_n$, to find a function of these, $t_n(x_1, x_2, \ldots, x_n)$, which should provide an estimate of an unknown parameter θ.

The desirable properties of estimation procedures have been discussed fully elsewhere. They are:

(*a*) That the estimator should be efficient according to some definition of efficiency previously arranged. Most commonly, the reciprocal of the variance of the estimates is taken as a measure of its efficiency, as this is most useful where central limit theory may be relevant.

(*b*) That the estimator should utilize all the information contained in the observations, $x_1, x_2, \ldots, x_n$ concerning the parameter θ. This is not always possible, but, if such an estimator exists, it is called sufficient.

(*c*) That the estimator should be consistent, i.e. t_n converges in some probabilistic sense to θ, usually $\lim_{n\to\infty} t_n \to \theta$.

(*d*) That the estimator should be unbiased, i.e. $E(t_n) = \theta$.

The method of maximum likelihood is popular in that it satisfies properties (*a*) to (*c*), whence, by evaluating $E(t_n)$, an unbiased statistic may be derived. That such evaluation is necessary is obvious when it is remembered that $\psi(t_n)$ is, by the same theory, the estimator of $\psi(\theta)$, and, since in general $E[\psi(t_n)] \neq \psi[E(t_n)]$, it will be the exception rather than the rule for a maximum-likelihood estimator to be unbiased.

Provided the exceptions may be simply evaluated no real difficulty arises. However, often the complexity of the evaluation presents a major drawback and some simple approach is then desirable.

2. If the observations are taken in random order, the estimator t_n may often be written

$$t_n = t_n(k_1, k_2, \ldots, k_m),$$

where $k_1, k_2, \ldots, k_m$ are unbiased estimates of the cumulants $\kappa_1, \kappa_2, \ldots, \kappa_m$. Then, provided that

$$\left.\begin{array}{ll} (a) & m \text{ is independent of } n, \\ (b) & \text{the function } t_n \text{ is capable of Taylorian expansion}, \\ (c) & \text{all of the cumulants are finite}, \\ (d) & t_n \text{ is consistent, i.e. } \theta = \lim_{n\to\infty} t_n(k_1 \ldots k_m), \end{array}\right\} \qquad \text{(I)}$$

it follows that

$$t_n - \theta = \Sigma(k_i - \kappa_i)\left(\frac{\partial t_n}{\partial k_i}\right)_{k_i=\kappa_i} + \Sigma\Sigma(k_i - \kappa_i)(k_j - \kappa_j)\left(\frac{\partial^2 t_n}{\partial k_i \partial k_j}\right)_{k_i=\kappa_i} + \ldots.$$

Since the moments of the estimators, k_i, are power series in $1/n$, it follows that $E(t_n - \theta)$, i.e. the bias in t_n, is also expressible as a power series in $1/n$.

The conditions (I) are undoubtedly more stringent than they need be. For instance, the higher cumulants need not exist. Further, it appears likely that I (*b*) is a necessary condition if I (*d*) is to hold. However, the main point is that for a wide variety of statistics it is true that

$$E(t_n-\theta)=\frac{a_1}{n}+\frac{a_2}{n^2}+\frac{a_3}{n^3}+\ldots.$$

3. If this is so, and $t'_n=nt_n-(n-1)t_{n-1}$, then

$$E(t'_n)=\theta-\frac{a_2}{n^2}-\frac{a_2+a_3}{n^3}-\ldots,$$

and hence t'_n is biased to order $1/n^2$ only. Similarly, $t''_n=[n^2t'_n-(n-1)^2t'_{n-1}]/[n^2-(n-1)^2]$ is biased to order $1/n^3$, and so on.

Alternatively, it is possible to use the statistics calculated from any subset of the observations to achieve corrections for bias. A further approach of particular interest occurs when $n=2p$. Here, we may use $t'_{2p}=2t_{2p}-t_p$ as being free from bias to order $1/n^2$.

Procedures such as these may supply approximate corrections for bias provided that efficiency of estimation is not lost in the process. To achieve this in general, it is necessary to use $\bar{t}_{n-1}$, the average of estimates from all possible sets of $n-1$ observations, instead of t_{n-1}, and similarly $\bar{t}_{n-2}$ instead of t_{n-2}, etc. With this provision, it appears likely that little, if any, loss of efficiency will result.

4. For instance, many of the statistics t_n may be derived from an estimation procedure of form

$$\sum_{i=1}^{n}G(x_i,t_n)=0.$$

The variance of t_n to the first order may be estimated from this equation by a δ technique such as has been described by Weatherburn (1952, pp. 130 et seq.) and Kendall (1943, pp. 208 et seq.). In the simplest instance it is possible to represent the argument as follows. If $\mu=E(x_i)$, then

$$n\delta t_n=H(\theta)\sum_{i=1}^{n}\delta x_i,$$

where

$$H(\theta)=E\left[\frac{\partial}{\partial\mu}G(\mu,\theta)\right]\Big/E\left[\frac{\partial}{\partial\theta}G(\mu,\theta)\right],$$

if both expectations exist. (If they do not exist, then generally the basic equation is changed to one of the form

$$n\delta t_n=\sum_{i=1}^{n}\delta H(x_i\theta),$$

to which a similar argument may be applied.)

Thus

$$\operatorname{var}t_n=\frac{[H(\theta)]^2}{n}\operatorname{var}x,$$

$$(n-1)\,\delta t_{n-1}=H(\theta)\sum_{i\neq j}\delta x_i,$$

$$(n-1)\,\delta\bar{t}_{n-1}=H(\theta)\frac{1}{n}\sum_{j=1}^{n}\sum_{i\neq j}\delta x_i$$

$$=H(\theta)\frac{n-1}{n}\sum_{i=1}^{n}\delta x_i.$$

Hence
$$\delta t'_n = n\delta t_n - (n-1)\,\delta \bar{t}_{n-1}$$
$$= \frac{H(\theta)}{n} \sum_{i=1}^{n} \delta x_i$$

and
$$\operatorname{var} t'_n = \frac{[H(\theta)]^2}{n} \operatorname{var} x = \operatorname{var} t_n$$

to order $1/n$.

This implies that this correction affects the standard error by a factor of $1/n$ at most, i.e.

$$\text{S.E. of } t'_n = (\text{S.E. of } t_n)\{1 + O(1/n)\}.$$

Since, in general, the standard error of t_n will decrease with n (usually proportional to $n^{-\frac{1}{2}}$), the correction will affect the dispersion of the distribution by $o(1/n)$ (usually, $O(n^{-\frac{3}{2}})$, i.e. by a small amount in comparison with the bias). The reduction in bias achieved by using t_n is consequently not obtained at the expense of a comparable increase in the dispersion of the distribution of the estimator.

5. As a first illustration, suppose $\theta = \sigma^2$ for a normal distribution, and $t_n = \sum_{i=1}^{n} (x_i - \bar{x})^2/n$. Then

$$t'_n = nt_n - (n-1)\,\bar{t}_{n-1}$$
$$= \frac{n \sum\sum_{i<j} (x_i - x_j)^2}{n^2} - \frac{n-1}{n} \sum_{k=1}^{n} \frac{\overset{\text{Excluding } k}{\sum\sum_{i<j}} (x_i - x_j)^2}{(n-1)^2}$$
$$= \sum\sum_{i<j} (x_i - x_j)^2 \left[\frac{1}{n} - \frac{n-2}{n(n-1)}\right]$$
$$= \frac{1}{n(n-1)} \sum\sum_{i<j} (x_i - x_j)^2$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

and
$$\operatorname{var} t'_n = \left(\frac{n}{n-1}\right)^2 \operatorname{var} t_n.$$

Similarly, if $t'_{2p} = 2t_{2p} - \bar{t}_p$, then
$$t'_{2p} = \frac{1}{2p-1} \sum_{i=1}^{2p} (x_i - \bar{x})^2$$

and
$$\operatorname{var} t'_{2p} = \left(\frac{2p}{2p-1}\right)^2 \operatorname{var} t_{2p} = \frac{2\sigma^4}{2p-1}.$$

If, alternatively, $\bar{t}_p$ is calculated from only one pair of possible sets of p observations, say x_1 to x_p and x_{p+1} to x_{2p}, then
$$t'_{2p} = \frac{1}{2p} \sum_{i=1}^{2p} x_i^2 - \frac{1}{p^2} \left(\sum_{i=1}^{p} x_i\right) \left(\sum_{i=p+1}^{2p} x_i\right)$$

and
$$\operatorname{var} t'_{2p} = \frac{p+1}{p^2}\,\sigma^4.$$

Thus averaging over only one pair of the possible sets results in a decrease in the efficiency of estimation of
$$\frac{(2p-1)(p+1)}{2p^2} - 1 = \frac{p-1}{2p^2}.$$

6. As a numerical illustration, suppose that it is desired to estimate $\theta = 1/\mu$ from a series of observations taken from a normal distribution. Then

$$t_n = n \Big/ \sum_{i=1}^{n} x_i$$

and

$$t'_n = \frac{n^2}{\sum_{i=1}^{n} x_i} - \frac{(n-1)^2}{n} \sum_{i=1}^{n} \frac{1}{\sum_{j \neq i} x_j},$$

$$\operatorname{var} t_n \sim \operatorname{var} t'_n \sim \frac{\sigma^2}{n\mu^4}.$$

The values in column 1 of Table 1 were random observations from a normal distribution with $\mu = 2$, $\sigma^2 = 1$ (using the first ten random numbers in Fisher & Yates's (1953) Statistical Tables).

Table 1

x_i	$18{\cdot}32 - x_i$	$t_{n-1} = 9/(18{\cdot}32 - x_i)$
0·18	18·14	0·4961
4·00	14·32	0·6285
1·04	17·28	0·5208
0·85	17·47	0·5152
2·14	16·18	0·5562
1·01	17·31	0·5199
3·01	15·31	0·5879
2·33	15·99	0·5629
1·57	16·75	0·5373
2·19	16·13	0·5580
18·32	$164{\cdot}88 = 9 \times 18{\cdot}32$	5·4828

Then

$$t_n = 1/1{\cdot}832 = 0{\cdot}54585,$$
$$t'_n = 5{\cdot}4585 - 9 \times 0{\cdot}54828 = 0{\cdot}5240.$$

Here, owing to the high value of s^2 ($= 1{\cdot}4$), this latter value has been corrected more than it would be using the exact formula

$$E(t_n) = \frac{n}{\sigma^2} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \int_0^{\mu} \exp\left(\frac{n\mu^2}{2\sigma^2}\right) d\mu \quad \text{(see appendix)}$$

$$= \frac{1}{\mu} + \frac{\sigma^2}{n\mu^3} + \ldots \quad \text{for large } \mu.$$

This might be compared with

$$t_{n-1} = \frac{n-1}{T - x_i} = \frac{n}{T}\left[1 + \frac{n(x_i - \bar{x})}{(n-1)T} + \frac{n^2(x_i - \bar{x})^2}{(n-1)^2 T^2} + \ldots\right], \quad \text{where} \quad T = \sum_{i=1}^{n} x_i,$$

$$\bar{t}_{n-1} = \frac{n}{T} + \frac{n^2 s^2}{(n-1)T^3} + \ldots,$$

$$nt_n - (n-1)\bar{t}_{n-1} = \frac{n}{T} - \frac{n^2 s^2}{T^3} - \ldots = \frac{1}{\bar{x}} - \frac{s^2}{n\bar{x}^3} - \ldots.$$

In this instance it might be noted that the procedure will probably break down if $n\mu^2/\sigma^2$ is small. This should be apparent from the behaviour of the t_{n-1}, which will vary in sign.

7. Consider next an inverse sampling procedure. Suppose the proportion, π, of individuals with a given characteristic is to be estimated, and sampling continues until r individuals with the characteristic are observed. Let n be the total number of individuals.

Let $t_n = r/n$, then since the last individual is constrained to have the characteristic, there are only $n-1$ values of t_{n-1} to be considered, and

$$\bar{t}_{n-1} = \frac{1}{n-1}\left[(r-1)\frac{r-1}{n-1} + (n-r)\frac{r}{n-1}\right] = \frac{nr-2r+1}{(n-1)^2}.$$

Thus

$$t'_n = n\frac{r}{n} - (n-1)\frac{nr-2r+1}{(n-1)^2} = \frac{r-1}{n-1},$$

which actually is strictly unbiased. Alternatively, if $1/\pi$ is to be estimated, then $t_n = n/r$, and

$$\bar{t}_{n-1} = \frac{1}{n-1}\left[(r-1)\frac{n-1}{r-1} + (n-r)\frac{n-1}{r}\right] = \frac{n}{r},$$

Thus $t'_n = t_n = n/r$, which again is strictly unbiased.

This indicates that the procedure may be useful in sequential estimation or inverse sampling.

8. It is possible to use simple extensions of this procedure to correct the bias in any combination of statistics, $f(t_n, u_m)$. The statistic

$$nmf(t_n, u_m) - (n-1)\,mf(t_{n-1}, u_m) - n(m-1)f(t_n, u_{m-1}) + (n-1)\,(m-1)\,f(t_{n-1}, u_{m-1})$$

is, for example, unbiased to order $1/n^2$ or $1/m^2$, whichever is the greater.

9. In general, where a series of concomitant observations, $y_1, y_2, \ldots, y_n$, are used in calculating t_n, both the bias and the efficiency will depend upon these observations, and therefore a simple correction of the above type will not be possible.

An important exception occurs when

$$t_n = \psi\left(\frac{\Sigma x_i \phi(y_i)}{\Sigma \phi^2(y_i)}\right). \tag{II}$$

In this instance, if

$$\xi = E[\Sigma x_i \phi(y_i)/\Sigma\phi^2(y_i)], \quad \sigma^2(t_n) = [\psi'(\xi)]^2 \operatorname{var} x / \sum_{i=1}^{n} \phi^2(y_i)$$

and

$$\frac{n-1}{\sigma^2(t_n)} = \Sigma \frac{1}{\sigma^2(t_{n-1})}.$$

Then, if $t'_n = nt_n - (n-1)\bar{t}_{n-1}$, where

$$\bar{t}_{n-1} = \left[\Sigma \frac{t_{n-1}}{\sigma^2(t_{n-1})} \Big/ \Sigma \frac{1}{\sigma^2(t_{n-1})}\right],$$

t'_n is unbiased to order $1/n^2$, and $\operatorname{var} t_n = \operatorname{var} t'_n$ to order $1/n^2$. This formula may thus be put in the alternative forms

$$t'_n = nt_n - \Sigma \frac{\sigma^2(t_n)}{\sigma^2(t_{n-1})} t_{n-1}$$

$$= nt_n - \Sigma\left(1 - \frac{\phi^2(y_n)}{\Sigma\phi^2(y_i)}\right) t_{n-1}.$$

In a similar manner, if

$$t_n = \psi\left(\frac{\Sigma x_i\{\phi(y_i)-\overline{\phi(y_i)}\}}{\Sigma\{\phi(y_i)-\overline{\phi(y_i)}\}^2}\right),$$

then

$$\frac{n-2}{\sigma^2(t_n)} = \frac{n-1}{n}\Sigma\frac{1}{\sigma^2(t_{n-1})}$$

and

$$t'_n = (n-1)\,t_n - (n-2)\,\bar{t}_{n-1},$$

where

$$\bar{t}_{n-1} = \left[\Sigma\frac{t_{n-1}}{\sigma^2(t_{n-1})}\right]\Big/\left[\Sigma\frac{1}{\sigma^2(t_{n-1})}\right]$$

is unbiased to order $1/n^2$.

It is also possible to split $2p$ observations into groups of p. Thus, if equation (II) holds,

$$\frac{1}{\sigma^2(t_{2p})} = \frac{1}{\sigma^2(t_{p,1})} + \frac{1}{\sigma^2(t_{p,2})},$$

and

$$t'_{2p} = 2t_{2p} - \bar{t}_p, \quad \text{where} \quad \bar{t}_p = \left(\Sigma\frac{t_p}{\sigma^2(t_p)}\right)\Big/\left(\Sigma\frac{1}{\sigma^2(t_p)}\right),$$

is unbiased to order $1/n^2$ and has equal asymptotic efficiency to t_{2p}.

If, however, a correction is made for the mean, some loss in efficiency will arise from the difference between the values of $\overline{\phi(y_i)}$ for the two groups. If these are equal, the above approach may be used with no extra loss in efficiency, and, since there is no ordering involved in statistics of this type, approximate equality of these values may often be achieved by appropriate selection. For instance, if the $\phi(y_i)$ are ordered and the pairs of observations corresponding to alternate values are used in the two groups, then both $\Sigma\phi(y_i)$ and $\Sigma\phi^2(y_i)$ will be approximately equal for the two groups and $t'_{2p} = 2t_{2p} - \frac{1}{2}(t_{p1}+t_{p2})$ will frequently be a sufficiently accurate and unbiased statistic.

None of these results is, however, of much practical importance in that the bias may be calculated directly using $\psi''(\xi)\,\sigma^2(t_n)/2[\psi'(\xi)]^2$. Their interest lies in that they indicate that the same corrections applied to time-series statistics may be adequate for many purposes. This has already been suggested elsewhere (Quenouille, 1949).

10. Consider first the application of these corrections in the estimation of a serial covariance, say the pth. Here, if the mean is known, say O, an unbiased estimator exists:

$$t_n = (x_1x_{p+1} + x_2x_{p+2} + \ldots + x_{n-p}x_n)/(n-p),$$

with variance $\sigma^4/(n-p)$ in the case where no correlation exists in the series, and variance $\sigma^4A/(n-p)$ otherwise, where A depends upon the correlations between the products in this expression.

There exist also n estimators based upon $n-1$ observations. Denoting the estimator which omits the ith observation by $t_{n-1,i}$, this has $n-p-1$ terms if $i \leqslant p$ or $\geqslant n-p+1$, and $n-p-2$ terms otherwise. The variance of $t_{n-1,i}$ is correspondingly $\sigma^4/(n-p-1)$ or $\sigma^4/(n-p-2)$ when no correlation exists in the series, and $\sigma^4A_i/(n-p-1)$ or $\sigma^4A_i/(n-p-2)$ if correlation exists. Here, A_i will differ for different i, but for large n it will approximate to A for all i.

If there is no serial correlation or if we ignore the differences in A_i, the analysis then gives

$$\begin{aligned} t_{n-1} &= \left(\Sigma\frac{t_{n-1}}{\sigma^2(t_{n-1,i})}\right)\Big/\left(\Sigma\frac{1}{\sigma^2(t_{n-1,i})}\right) \\ &= (n-2)\,(n-p)\,t_n/[n(n-p-2)+2p] = t_n, \\ t'_n &= t_n. \end{aligned}$$

Thus the correction does not affect the estimate.

If the variation in A_i is taken into account, a slightly different estimate is obtained. This is still unbiased, but is less efficient (though asymptotically of equal efficiency) probably as a consequence of the individual estimators not being fully efficient. For example, the products in t_n are correlated with one another, and hence the end products contain information on $x_{n-p+1}x_{n+1}$, etc. The end-products should thus receive slightly greater weight for efficient estimation. Similarly, some of the products in $t_{n-1\,i}$ should receive greater weight. With these provisions, it appears likely that the above procedure would not lead to any loss in efficiency, i.e. the slight loss in efficiency (asymptotically zero) which occurs results from the use of inefficient estimates. This obviously is of no practical importance.

The effectiveness of corrections of this type in the general case is more difficult to prove. Their effectiveness might be demonstrated by considering the correction for the mean in the estimation of serial covariance.

If

$$t_n = \frac{1}{n-p}\left[\sum_{i=1}^{n-p} x_i x_{i+p} - \frac{\left(\sum_{i=1}^{n-p} x_i\right)\left(\sum_{i=1}^{n-p} x_{i+p}\right)}{n-p}\right],$$

then, when the x_i are uncorrelated,

$$E(t_n) = -\frac{n-2p}{(n-p)^2}\sigma^2 = -\frac{\sigma^2}{n}\left[1+O\left(\frac{1}{n^2}\right)\right].$$

The expectations of the t_{n-1} will vary. For instance, there will be a few terms of the type

$$E(t_{n-1,1}) = -\frac{E\left(\sum_{i=2}^{n-p} x_i\right)\left(\sum_{i=2}^{n-p} x_{i+p}\right)}{(n-p-1)^2} = -\frac{n-2p-1}{(n-p-1)^2}\sigma^2 = -\frac{\sigma^2}{n-1}\left[1+O\left(\frac{1}{n^2}\right)\right],$$

and

$$E(t_{n-1,2}) = -\frac{E\left(x_1+\sum_{i=3}^{n-p} x_i\right)\left(x_{p+1}+\sum_{i=3}^{n-p} x_{i+p}\right)}{(n-p-1)^2} = -\frac{\sigma^2}{n-1}\left[1+O\left(\frac{1}{n^2}\right)\right] \quad (p \neq 1),$$

but the majority ($n-4p$ out of n) will be of the form

$$E(t_{n-1,m}) = -\frac{E\left(\sum_{i=1}^{n-p} x_i - x_{m-p} - x_m\right)\left(\sum_{i=1}^{n-p} x_{i+p} - x_m - x_{m+p}\right)}{(n-p-2)^2}$$

$$= -\frac{n-2p-3}{(n-p-2)^2}\sigma^2 = -\frac{\sigma^2}{n-1}\left[1+O\left(\frac{1}{n^2}\right)\right].$$

Thus

$$E(\bar{t}_{n-1}) = -\frac{\sigma^2}{n-1}\left[1+O\left(\frac{1}{n^2}\right)\right] \quad \text{and} \quad E(t'_n) = O\left(\frac{1}{n^2}\right)$$

as required.

Similar results will hold for the serial correlation coefficients. It is, however, an open question as to whether the extra computation involved in calculating and using $\bar{t}_{n-1}$ is warranted compared with that involved in calculating and using $t_{(\frac{1}{2}n)}$. It appears likely that the use of the two half-series is sufficiently accurate for most practical purposes though this point requires further investigation.

REFERENCES

FISHER, R. A. & YATES, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh: Oliver and Boyd.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Griffin and Co.

QUENOUILLE, M. H. (1949). *J. R. Statist. Soc.* B, **11**, 68–84.

WEATHERBURN, C. E. (1952). *Mathematical Statistics.* Cambridge University Press.

APPENDIX

Proof of a formula in § 6

$$E(t_n) = \int_{-\infty}^{\infty} \frac{1}{x} \frac{\sqrt{n}}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{n(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{\sqrt{n}}{\sigma} \int_{-\infty}^{\infty} \frac{1}{y\sqrt{(2\pi)}} \exp\{-\tfrac{1}{2}(y-a)^2\}\, dy,$$

where

$$a = \mu\sqrt{n}/\sigma,\ y = x\sqrt{n}/\sigma.$$

Let

$$I(a) = \int_{-\infty}^{\infty} \frac{1}{y\sqrt{(2\pi)}} \exp\{-\tfrac{1}{2}(y-a)^2\}\, dy,$$

then

$$\exp(\tfrac{1}{2}a^2)\, I(a) = \int_{-\infty}^{\infty} \frac{1}{y\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}y^2 + ay)\, dy,$$

$$\frac{\partial}{\partial a}[\exp(\tfrac{1}{2}a^2)\, I(a)] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}y^2 + ay)\, dy$$

$$= \exp(\tfrac{1}{2}a^2).$$

Thus

$$\exp(\tfrac{1}{2}a^2)\, I(a) = \int_0^a \exp(\tfrac{1}{2}a^2)\, da,$$

the limits being determined by the fact that $I(a) = 0$ when $a = 0$.

Therefore

$$I(a) = \exp(-\tfrac{1}{2}a^2) \int_0^a \exp(\tfrac{1}{2}a^2)\, da$$

and

$$E(t_n) = \frac{n}{\sigma^2} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \int_0^{\mu} \exp\left(\frac{n\mu^2}{2\sigma^2}\right) d\mu.$$

AN INTRODUCTION TO SOME NON-PARAMETRIC GENERALIZATIONS OF ANALYSIS OF VARIANCE AND MULTIVARIATE ANALYSIS*

By S. N. ROY and S. K. MITRA

Institute of Statistics, University of North Carolina

It is clear that a p-variate body of data arranged in a q-way classification will formally look like a $(p+q)$-dimensional contingency table, but a distinction can be made between a 'variate' and a 'way of classification' in that along the direction of a 'variate' the marginal frequencies are supposed to be stochastic variates while along a 'way of classification' the marginal frequencies are supposed to be fixed. When, along certain directions, the marginals are fixed, an approach based on a conditional probability argument has been used.

In the present paper (i) the conditional probability approach is abandoned and we start either from a single multinomial distribution or a product of an appropriate number of different multinomial distributions according as, with multi-way frequency data, all ways are 'variates' or some are 'variates' and some are 'ways of classification'. (ii) Also the hypotheses that are posed are of different kinds altogether according as we have a 'multivariate analysis' situation or an 'analysis of variance' situation. The hypotheses that are meaningful for one situation would not be too meaningful for the other and vice versa. Since the conditional probability approach is altogether abandoned, the mathematical theorems to which appeal is made are the two theorems as stated and proved by Cramér (1946, chapter 30) and a number of other such theorems which have been proved the same way and which, between them, take care of all the hypotheses discussed in this paper. When all ways are 'variates' the hypotheses are analogous to those in the usual multivariate analysis, and when some ways are 'variates' and some are 'ways of classification' the hypotheses are analogous to those in the analysis of variance.

The general methods discussed in this paper arose out of an attempt to analyse a large mass of categorical data. The analysis has been carried out, and a few typical cases (together with the numerical analysis) illustrating different parts of the theoretical development will be presented in a subsequent paper. In this paper only the large-sample tests are considered. How 'large' the sample size has to be for the validity of the use of these asymptotic techniques, or in other words, some results on the nature of the approximation involved will also be discussed in a later paper.

1. Introduction

After this paper had been written the attention of the authors was drawn to the papers by Barnard (1947) and Pearson (1947), which the authors have since studied. Although it is only the 2×2 table with which Barnard is exclusively and Pearson is mostly concerned, their logical approach to the whole problem is exactly the same as the authors, and they have stated the whole position with admirable lucidity. The authors feel that in this general area, aside from the pioneering work of Karl Pearson, of Fisher on χ^2 (1922) and of Yule & Kendall (1950) on association, the four conspicuous landmarks in recent times must be Cramér's chapter on χ^2 (1946, chapter 30), Neyman's paper in the same area (1949) and the two papers by Barnard and Pearson just referred to. In historical perspective, the present paper might, therefore, be regarded as an attempt at a somewhat systematic elaboration and application of the earlier ideas of Barnard and Pearson (i) to (mostly) one population multivariate analysis, starting from a single multinomial distribution and framing hypotheses appropriate to multivariate analysis situations, and (ii) to analysis of variance, starting from the product of an appropriate number of separate multinomial distributions and framing hypotheses appropriate to analysis of variance situations.

* This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

The authors (presumably for the same reasons as Barnard and Pearson) abandon the conditional probability approach for a physical and a mathematical reason. The physical reason is that keeping the marginals fixed in the sense of conditional probability is seldom experimentally realizable, and although the authors are not dogmatic about this point, they would prefer to keep their probability statements as close as possible to experimentally realizable processes. The mathematical reason is the following. Let n_i $(i = 1, 2, \ldots, r)$ be the observed frequency in the ith cell and p_i be the probability (under any hypothesis) of getting an observation in the ith cell, and let the observations be independent in probability. Let $\sum_i n_i = n$ be supposed to be fixed, $p_i > 0$ and $\sum_i p_i = 1$. Then the unconditional likelihood function is $n! \prod_i p_i^{n_i} / \prod_i n_i!$. Next let us denote by $\mathbf{n}'$ the row vector $(n_1, \ldots, n_r)$ and let A be an $s \times r$ $(s < r)$ matrix of constants of rank, say $t \leqslant s$, and let us suppose that the n_i's are subject to the constraints $A\mathbf{n} = \mathbf{a}^0$, a constant vector. Then the conditional likelihood function will be given by

$$\phi = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i} \Big/ \sum_{n:\, A\mathbf{n} = \mathbf{a}^0} \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}. \tag{1·1}$$

Assume, for simplicity of discussion, that the p_i's are completely specified by the hypothesis (although this makes no essential difference to the argument). Is $\sum_{i=1}^{r} (n_i - np_i)^2/(np_i)$, in large samples, distributed as χ^2 with degrees of freedom $(r-1)-t$, no matter what A might be, provided that it is of rank t? If the answer to this question is yes, then the customary linear estimation or testing of linear hypotheses (in the sense of least squares and analysis of variance) can be carried out, starting from (1·1) and eventually using the χ^2-test. One of the authors (Mitra) has constructed an A matrix such that under a ϕ (of 1·1) with that A matrix, the limiting distribution of $\sum_i (n_i - np_i)^2/(np_i)$ is not a χ^2-distribution. This, of course, will not affect the validity of the applications, mostly simple, actually made over the last fifty years starting from (1·1), because, in all these applications, A is such that we have the χ^2-distribution. But it would be mathematically unsafe to try to set up a *general* theory of testing of linear hypotheses, starting from (1·1), to say nothing of its being physically unsatisfactory. We do not know under what restrictions on A the distribution will be χ^2, but the authors have shown that at least those A's are permissible under which, instead of starting from (1·1), we could also have abandoned the conditional probability approach and started from the product of a number of different multinomial distributions. The invalidating example will be discussed in a later paper.

2. Problems in a two-way table

To fix our ideas, consider first a two-way, say $r \times s$, table with observed frequencies n_{ij} in the (ij)th cell $(i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, s)$. Also let

$$\sum_i n_{ij} = n_{0j}, \quad \sum_j n_{ij} = n_{i0} \quad \text{and} \quad \sum_{i,j} n_{ij} = n_{00} = n \quad \text{(say)}.$$

2·1. *Both 'i' and 'j' are 'variates'*

Assume that we have a sample of n independent observations such that p_{ij} is the probability of an observation in the (ij)th cell, and n is fixed from sample to

sample. Also let $\sum_j p_{ij} = p_{i0}$, $\sum_i p_{ij} = p_{0j}$, $\sum_{i,j} p_{ij} = p_{00} = 1$. Then the likelihood function is given by

$$\phi = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{i,j} p_{ij}^{n_{ij}}. \tag{2·1}$$

The composite hypothesis we shall be interested in testing is that 'i' and 'j' are independent, that is, H_0: $p_{ij} = p_{i0}p_{0j}$ against $H \neq H_0$, where the p_{i0}'s and p_{0j}'s are arbitrary positive nuisance parameters subject to $\sum_i p_{i0} = \sum_j p_{0j} = 1$. This is the analogue of the hypothesis of no correlation in a bivariate normal population. Under H_0 we shall have the likelihood function ϕ_0 given by

$$\phi_0 = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{i,j} (p_{i0}p_{0j})^{n_{ij}} = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_i p_{i0}^{n_{i0}} \prod_j p_{0j}^{n_{0j}}. \tag{2·1·1}$$

2·2. *'i' is a 'way of classification' and 'j' is a 'variate'*

Assume that we have r independent sets of sizes $n_{10}, n_{20}, \ldots, n_{r0}$ of independent observations such that n_{i0} $(i = 1, 2, \ldots, r)$ is fixed from sample to sample and p_{ij} is the probability of an observation in the (ij)th cell. Also we notice that $\sum_j p_{ij} = p_{i0} = 1$. Then the likelihood function is given by

$$\phi = \prod_i \left[\frac{n_{i0}!}{\prod_j n_{ij}!} \prod_j p_{ij}^{n_{ij}} \right]. \tag{2·2}$$

The composite hypothesis we shall be interested in testing is that p_{ij}, for any j, is independent of 'i', or in other words, H_0: $p_{ij} = q_{0j}$ (say) against $H \neq H_0$, where the q_{0j}'s are arbitrary positive nuisance parameters subject to $\sum_j q_{0j} = \sum_j p_{ij} = p_{i0} = 1$. This is the analogue of the hypothesis of the equality of means for r homoscedastic univariate normal populations. Under H_0 we shall have

$$\phi_0 = \prod_i \left[\frac{n_{i0}!}{\prod_j n_{ij}!} \prod_j q_{0j}^{n_{ij}} \right] = \frac{\prod_i n_{i0}!}{\prod_{i,j} n_{ij}!} \prod_j q_{0j}^{n_{0j}}. \tag{2·2·1}$$

This ϕ_0 could also have been obtained from the ϕ of (2·1), by putting H_0: $p_{ij} = p_{i0}p_{0j}$ and then finding under H_0 the conditional probability subject to the n_{i0}'s being fixed. But it seems to the authors that physically this is far less realistic than the model here used, although historically this is more or less what has been used so far.

The case of 'i' being a 'variate' and 'j' a 'way of classification' is exactly similar and need not be separately considered.

2·3. *Both 'i' and 'j' are 'ways of classification'*

Here we have a sampling scheme in which the n_{i0}'s and n_{0j}'s $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, s)$ are supposed to be fixed from sample to sample. In this situation, on the hypothesis of independence between 'i' and 'j', we can write down the likelihood function ϕ_0 without assuming that the observations are independent. For this we start from an urn problem model in which there is an urn containing $n_{10}, n_{20}, \ldots, n_{r0}$ balls of r different colours from which we draw successively without replacement $n_{01}, n_{02}, \ldots, n_{0s}$ balls (with $\sum_i n_{i0} = \sum_j n_{0j} = n$).

The joint probability that the jth set of n_{0j} balls will contain $n_{1j}, n_{2j}, \ldots, n_{rj}$ balls of different colours (with $j = 1, 2, \ldots, s$) will be given by

$$\phi_0 = \prod_i n_{i0}! \prod_j n_{0j}!/(n! \prod_{i,j} n_{ij}!). \tag{2·3}$$

The great advantage of this scheme is that the different observations need not be assumed to be independent, and the great disadvantage is that it is not clear what is the form of ϕ under a general H as distinct from the null hypothesis H_0 of independence between 'i' and 'j'. This means that the power of a test for H_0 against alternatives cannot be obtained and also that is is not possible to obtain a one-tailed χ^2-test for H_0 using the same kind of heuristic arguments that we shall use in the first two situations. A one-tailed χ^2-test can be found here by analogy with what is done in the first two cases.

This ϕ_0 could also have been obtained from the ϕ of (2·1), by putting H_0: $p_{ij} = p_{i0}p_{0j}$ and then finding under H_0 the conditional probability subject to the n_{i0}'s and n_{0j}'s being fixed. But this would be less realistic than the model here used and would deprive (2·3) of the one great advantage it possesses in that the successive observations do not have to be independent. Notice that (2·1) is based on the observations being all independent.

It will be seen that the approach here is not one of conditional probability at all. It will also be seen that there are three different sampling schemes each leading in a natural way to a particular probability model and a particular type of hypothesis to be tested. From a physical standpoint it would not be proper to break this tie and use a particular probability model and test a particular type of hypothesis when the sampling scheme is something different. It will be noticed that in most situations of life the natural sampling schemes are those of (i) or (ii), but there are situations, e.g. Fisher's tea-tasting experiments or those connected with the extra-sensory perception experiments or with the claims of astrologers as to prediction, etc., where (iii) might be a natural sampling scheme.

3. Problems in a three-way table

As a natural extension of a two-way table consider a three-way $r \times s \times t$ table with observed frequencies n_{ijk} in the (ijk)th cell ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$). Also let

$$\sum_i n_{ijk} = n_{0jk}, \quad \sum_j n_{ijk} = n_{i0k}, \quad \sum_k n_{ijk} = n_{ij0}, \quad \sum_{i,j} n_{ijk} = n_{00k},$$

$$\sum_{i,k} n_{ijk} = n_{0j0}, \quad \sum_{j,k} n_{ijk} = n_{i00}, \quad \sum_{i,j,k} n_{ijk} = n_{000} = n \quad \text{(say)}.$$

3·1. 'i', 'j' and 'k' all '*variates*'

Assume that we have a sample of n independent observations such that p_{ijk} is the probability of an observation in the (ijk)th cell and n is fixed from sample to sample. Also let

$$\sum_i p_{ijk} = p_{0jk}, \quad \sum_j p_{ijk} = p_{i0k}, \quad \sum_k p_{ijk} = p_{ij0}, \quad \sum_{i,j} p_{ijk} = p_{00k},$$

$$\sum_{i,k} p_{ijk} = p_{0j0}, \quad \sum_{j,k} p_{ijk} = p_{i00}, \quad \sum_{i,j,k} p_{ijk} = p_{000} = 1.$$

The likelihood function will be given by

$$\phi = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}}. \tag{3·1}$$

In this case we shall be interested in testing a class of composite hypotheses, a typical one being the composite

3·1*a*. *Hypothesis of conditional independence between 'i' and 'j' for fixed 'k'*

This will be

$$H_0: \quad \frac{p_{ijk}}{p_{00k}} = \frac{p_{i0k}}{p_{00k}} \frac{p_{0jk}}{p_{00k}} \quad \text{or} \quad p_{ijk} = \frac{p_{i0k} p_{0jk}}{p_{00k}}, \tag{3·1·1}$$

against $H \neq H_0$ ($i = 1, \ldots, r$; $j = 1, \ldots, s$; $k = 1, \ldots, t$).

This is the analogue of the hypothesis of no partial correlation between x and y, given z, in a three-variate normal population. It is easy to see that if we superimpose on this the composite hypothesis of independence between 'i' and 'k', and between 'j' and 'k', i.e.

$$p_{i0k} = p_{i00} p_{00k} \quad \text{and} \quad p_{0jk} = p_{0j0} p_{00k}, \tag{3·1·2}$$

which is the analogue of the hypothesis of no total correlation between x and z, and between y and z, in a three-variate normal population, we should have

$$p_{ijk} = p_{i00} p_{0j0} p_{00k}, \tag{3·1·3}$$

which is the condition of complete independence of 'i', 'j' and 'k'.

We shall also be interested in another class of composite hypotheses, the typical one being the composite

3·1*b*. *Hypothesis of independence between '(i,j)' and 'k'*

This will be

$$H_0: \quad p_{ijk} = p_{ij0} p_{00k} \text{ against } H \neq H_0 \quad (i = 1, \ldots, r;\ j = 1, \ldots, s;\ k = 1, \ldots, t). \tag{3·1·4}$$

This is the analogue of the hypothesis of no multiple correlation between (x, y) and z in a three-variate normal population.

It is easy to check by summing the two sides of the above equation over i and also over j separately that (3·1·4) implies the composite hypotheses

$$p_{i0k} = p_{i00} p_{00k} \quad \text{and} \quad p_{0jk} = p_{0j0} p_{00k}. \tag{3·1·5}$$

But (3·1·5) will not imply (3·1·4). It has been shown by Roy & Kastenbaum (1956) that it is necessary to superimpose on (3·1·5) the additional hypothesis

$$H_1: \quad p_{ijk} = q_{ij0} q_{i0k} q_{0jk}, \tag{3·1·6}$$

to obtain (3·1·4), where q_{ij0}, q_{i0k}, q_{0jk} are defined to be arbitrary positive functions of (i,j), (i,k) and (j,k) with no summation convention connecting them as in the case of the p_{ijk}'s. By analogy with analysis of variance H_1 will be called the hypothesis of 'no interaction'.

3·2. *'i' and 'j' are 'variates' and 'k' a 'way of classification'*

Assume that we have t independent sets of sizes $n_{001}, \ldots, n_{00t}$ of independent observations such that n_{00k} ($k = 1, \ldots, t$) is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)th cell. Notice that $\sum_{i,j} p_{ijk} = p_{00k} = 1$. The likelihood function will be given by

$$\phi = \prod_k \left[\frac{n_{00k}!}{\prod_{i,j} n_{ijk}!} \prod_{i,j} p_{ijk}^{n_{ijk}} \right]. \tag{3·2}$$

Here we shall be interested in testing the composite

3·2*a*. *Hypothesis of independence between 'i' and 'j' for each 'k'*, that is,

$$H_0: \quad p_{ijk} = p_{i0k} p_{0jk} \text{ against } H \neq H_0 \quad (i = 1, \ldots, r;\ j = 1, \ldots, s;\ k = 1, \ldots, t). \tag{3·2·1}$$

If we superimpose on this the composite hypothesis that the marginal 'i' (obtained by summing over 'j') is independent of 'k' and similarly for 'j', that is,

$$p_{i0k} = q_{i00} \text{ (say)} \quad \text{and} \quad p_{0jk} = q_{0j0} \text{ (say)}, \tag{3·2·2}$$

we should have

$$p_{ijk} = q_{i00}q_{0j0}. \tag{3·2·3}$$

Notice from (3·2·2) that $\sum_i q_{i00} = \sum_i p_{i0k} = p_{00k} = 1$ and also that $\sum_j q_{0j0} = \sum_j p_{0jk} = p_{00k} = 1$.

We shall also be interested in the composite

3·2*b*. *Hypothesis that p_{ijk} is independent of 'k'*, that is,

$$H_0: \quad p_{ijk} = q_{ij0} \text{ (say) against } H \neq H_0 \quad \text{(for all } i, j \text{ and } k). \tag{3·2·4}$$

This is the analogue of the hypothesis of the equality of t mean vectors (each consisting of two components) for t bivariate normal populations, each having the same variance-covariance matrix.

Summing over 'j' and 'i' separately this would imply

$$p_{i0k} = \sum_j q_{ij0} = q_{i00} \text{ (say)} \quad \text{and} \quad p_{0jk} = \sum_i q_{ij0} = q_{0j0} \text{ (say)}. \tag{3·2·5}$$

As in the case where 'i', 'j' and 'k' are all 'variates', so also here, (3·2·4) implies (3·2·5) but (3·2·5) does not imply (3·2·4). Exactly in the same way as shown by Roy & Kastenbaum (1956) it can be shown that the extra condition which, when superimposed on (3·2·5), will imply (3·2·4) is

$$p_{ijk} = q_{ij0}q_{0jk}q_{i0k}. \tag{3·2·6}$$

3·3. *'i' is a 'variate' and 'j' and 'k' are 'ways of classification'*

Assume that we have $s \times t$ independent sets of sizes n_{0jk} of independent observations such that n_{0jk} ($j = 1, \ldots, s$; $k = 1, \ldots, t$) is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)th cell. Notice that $\sum_i p_{ijk} = p_{0jk} = 1$. The likelihood function will be given by

$$\phi = \prod_{j,k}\left[\frac{n_{0jk}!}{\prod_i n_{ijk}!}\prod_i p_{ijk}^{n_{ijk}}\right]. \tag{3·3}$$

Here we shall be interested in the composite

3·3*a*. *Hypothesis that for any 'k', p_{ijk} is independent of 'j'*, that is,

$$H_0: \quad p_{ijk} = q_{i0k} \text{ (say) against } H \neq H_0 \quad \text{(for all } i, j \text{ and } k). \tag{3·3·1}$$

Notice that $\sum_i q_{i0k} = \sum_i p_{ijk} = p_{0jk} = 1$.

We shall be also interested in the other composite

3·3*b*. *Hypothesis that for any 'j', p_{ijk} is independent of 'k'*, that is,

$$H_0: \quad p_{ijk} = q_{ij0} \text{ (say), against } H \neq H_0 \quad \text{(for all } i, j \text{ and } k). \tag{3·3·2}$$

Notice that $\sum_i q_{ij0} = \sum_i p_{ijk} = p_{0jk} = 1$. We now observe that (3·3·1) together with (3·3·2) implies that p_{ijk} is a pure function of 'i', i.e. that

$$p_{ijk} = q_{i00} \text{ (say)} \quad \text{(for all } i, j \text{ and } k). \tag{3·3·3}$$

If, in a one-way classification in the usual analysis of variance, 'i' corresponds to the 'variate', 'j' to the 'concomitant variate' and 'k' to the 'way of classification', then it will be seen on a little reflexion that (3·2·1) will be the analogue of the hypothesis of no regression and (3·2·4) will be the analogue of the hypothesis of no covariance. On the other hand, suppose we take 'j' and 'k' as just two 'ways of classification', for example, if we take 'j' as, say, blocks and 'k' as, say, treatments in a randomized block experiment (with more than one and in general unequal number of replications in each cell). Then (3·3·1) will be the analogue of 'no block effect' for each treatment separately and (3·3·2) will be the analogue of 'no treatment effect' for each block separately. In other words, in the usual parlance of analysis of variance, (3·3·1) combines the hypothesis of 'no main effect' and 'no interaction', while (3·3·2) combines the hypotheses of another 'no main effect' and 'no interaction'. In the analysis of variance situations for data which are not of the 'normal variate' type, the authors believe that this would be a better way of handling the material than the one used by Roy & Kastenbaum. Even for 'normal variate' type of data in the analysis of variance situations the authors are not sure that from the physical standpoint this might not be a better approach than the customary one of analysis into main effects and interactions of various orders. It is hoped to consider this in a later paper.

3·4. *'i' is a variate and 'j' and 'k' are 'ways of classification' in the sense of a 'balanced incomplete' or 'partially balanced incomplete' or a more general type of 'incomplete' block experiment*

Assume as before that there are r 'i''s, s 'j''s and t 'k''s. Assume further that 'j' is a block and 'k' a treatment and that, for any 'j', there is a set of treatments $(T)_j$ associated with it, of number t_j. In other words, for a given j, k takes on the set of values $(t)_j$, where $(t)_j$ is a set of indices of number t_j out of $1, 2, \ldots, t$. Now assume that we have $\sum_{j=1}^{s} t_j$ independent sets of sizes n_{0jk} of independent observations such that n_{0jk} $(k \in (t)_j;\ j = 1, 2, \ldots, s)$ is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)th cell. As before $\sum_i p_{ijk} = p_{0jk} = 1$. The likelihood function will be given by

$$\phi = \prod_{j=1}^{s} \prod_{k \in (t)_j} \left[\frac{n_{0jk}!}{\prod_i n_{ijk}!} \prod_i p_{ijk}^{n_{ijk}} \right]. \tag{3·4}$$

We can take over the hypothesis (3·3·1) of 'no block effect for each treatment separately' and (3·3·2) of 'no treatment effect for each block separately'. For a 'balanced incomplete design' all the t_j's will be equal and there will be a highly symmetrical pattern while for a 'partially balanced design' all the t_j's will be equal but there will be a less symmetrical pattern.

3·5. *'i' is a 'way of classification' and 'j, k' also are 'ways of classification' in the sense that the n_{i00}'s and n_{0jk}'s are fixed from sample to sample*

We can write down ϕ_0 in this case in exactly the same way as we wrote down the ϕ_0 in § 2·3. On the hypothesis of independence between 'i' and '(j, k)', this will be

$$\phi_0 = \prod_i n_{i0}! \prod_{j,k} n_{0jk}! / (n! \prod_{i,j,k} n_{ijk}!). \tag{3·5}$$

Starting from this we can test the hypothesis of independence between 'i' and 'j, k'.

The case of 'i', 'j' and 'k' being 'ways of classification' in the sense of the n_{i00}'s, n_{0j0}'s and n_{00k}'s (but not the n_{0jk}'s) being fixed from sample to sample is also of some interest, but we shall not consider that case in the present paper.

The extension of the problems of the two-way tables of § 2 to those of the three-way tables of § 3 is a rather big conceptual jump, but the extension from three-way tables to those of higher dimensions involves no such jump and will not be discussed in this paper except for some remarks towards the end.

4. The derivation of the χ^2 test on the union-intersection principle

Let a random sample of size n from some population be classified into k ($<n$) mutually exclusive and exhaustive categories according to some observable characteristics (qualitative or quantitative), and let the probability of a random observation falling in the ith category be p_i with $p_i>0$ and $\sum_{i=1}^{k} p_i = 1$. Let n_i denote the observed frequency in the ith category with, of course, $\sum_i n_i = n$. Also let $\mathbf{n}' = (n_1, n_2, \ldots, n_k)$ and $\mathbf{p}' = (p_1, p_2, \ldots, p_k)$. We have now

$$P[\mathbf{n}' \mid \mathbf{p}'] = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}. \tag{4}$$

4·1. *A simple hypothesis H_0: $\mathbf{p}' = \mathbf{p}_0'$ against the composite alternative H: $\mathbf{p}' \neq \mathbf{p}_0'$*

Consider first the most powerful test at a level say $\beta_{\mathbf{p}_1}$ of H_0: $\mathbf{p}' = \mathbf{p}_0'$ against a *specific* $\mathbf{p}_1' \neq \mathbf{p}_0'$, which, by the Neyman–Pearson lemma, will be as follows:

Reject H_0 if

$$P[\mathbf{n}' \mid \mathbf{p}_1']/P[\mathbf{n}' \mid \mathbf{p}_0'] \geqslant \mu \tag{4·1·1}$$

and accept H_0 otherwise, where, given μ, the size of the critical region (4·1·1) under $\mathbf{p}_0'$ should be $\beta(\mu, \mathbf{p}_0', \mathbf{p}_1', n)$. Substituting in (4·1·1) from (4) and taking logarithms on both sides, we see after a little simplification that (4·1·1) becomes

$$\frac{\mathbf{a}'(\mathbf{p}_1)(\mathbf{n} - n\mathbf{p}_0)}{\sqrt{\{\mathbf{a}'(\mathbf{p}_1)\,\Lambda^0\mathbf{a}(\mathbf{p}_1)\}}} \geqslant \frac{\log\mu - n\mathbf{a}'(\mathbf{p}_1)\,\mathbf{p}_0}{\sqrt{\{\mathbf{a}'(\mathbf{p}_1)\,\Lambda^0\mathbf{a}(\mathbf{p}_1)\}}} = c(\mathbf{p}_1', \beta_{\mathbf{p}_1}, n) \quad \text{say,} \tag{4·1·2}$$

where $\mathbf{a}'(\mathbf{p}_1) = [\log(p_{11}/p_{10}),\ \log(p_{21}/p_{20}),\ \ldots,\ \log(p_{k1}/p_{k0})]$, and $\Lambda^0 = (\sigma_{ij}^0)$ is the variance covariance matrix of $n_1, n_2, \ldots, n_k$ under H_0: $\mathbf{p}' = \mathbf{p}_0'$ and where $\beta_{\mathbf{p}_1}$ is supposed to vary with, that is depend upon, $\mathbf{p}_1$. It is thus evident that, for a fixed c, the critical region

$$w(\mathbf{p}_1', c) = \left\{\mathbf{n}' : \frac{\mathbf{a}'(\mathbf{p}_1)[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\{\mathbf{a}'(\mathbf{p}_1)\,\Lambda^0\mathbf{a}(\mathbf{p}_1)\}}} \geqslant c\right\} \tag{4·1·3}$$

is the most powerful critical region for testing $\mathbf{p}' = \mathbf{p}_0'$ against a specific $\mathbf{p} = \mathbf{p}_1$ ($\neq \mathbf{p}_0'$) at a level of significance $\beta(\mathbf{p}_1', c, n)$. Since the composite H: $\mathbf{p}' \neq \mathbf{p}_0'$ is the union of all $\mathbf{p}_1' \neq \mathbf{p}_0'$, we use the union-intersection principle (Roy, 1953) and take for H_0 against the composite H the critical region

$$w(c) = \bigcup_{\mathbf{p}' \neq \mathbf{p}_0'} w(\mathbf{p}', c). \tag{4·1·4}$$

Thus we should have for the complement of $w(c)$

$$\left\{\mathbf{n} : \sup_{\mathbf{p}' \neq \mathbf{p}_0'} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\{\mathbf{a}'(\mathbf{p})\,\Lambda^0\mathbf{a}(\mathbf{p})\}}} < c\right\}. \tag{4·1·5}$$

Since $\sum_i n_i = n$ and $\sum_i p_{i0} = 1$, we can write

$$\frac{\mathbf{a}'(\mathbf{p})[\mathbf{n}-n\mathbf{p}_0]}{\sqrt{\{\mathbf{a}'(\mathbf{p})\Lambda^0\mathbf{a}(\mathbf{p})\}}} = \frac{\mathbf{b}'(\mathbf{p})[\mathbf{n}^*-n\mathbf{p}_0^*]}{\sqrt{\{\mathbf{b}'(\mathbf{p})\Lambda^0_{kk}\mathbf{b}(\mathbf{p})\}}}, \tag{4·1·6}$$

where $\mathbf{n}^{*\prime} = (n_1, n_2, \ldots, n_{k-1})$, $\mathbf{p}_0^{*\prime} = (p_{10}, p_{20}, \ldots, p_{k-1,0})$,

$$\mathbf{b}'(\mathbf{p}) = (b_1(\mathbf{p}), b_2(\mathbf{p}), \ldots, b_{k-1}(\mathbf{p})),$$

$$b_i(\mathbf{p}) = a_i(\mathbf{p}) - a_k(\mathbf{p}) = \log(p_i p_{k0}/p_k p_{i0}) \quad (i = 1, \ldots, k-1)$$

and Λ^0_{kk} is the matrix formed by omitting the kth row and the kth column of Λ^0. Notice that each $b_i(\mathbf{p})$ can assume any value on the real line and conversely, given any real vector $\mathbf{b}_0' = (b_{10}, \ldots, b_{k-1,0})$, the equations $\mathbf{b}'(\mathbf{p}) = \mathbf{b}_0'$ have always a unique solution in $\mathbf{p}$; thus

$$p_i/p_k = (p_{i0}/p_{k0})e^{b_{i0}} = \lambda_{i0}, \quad \text{say},$$

or $\quad p_i = \lambda_{i0}\Big/\Big(1+\sum_{i=1}^{k-1}\lambda_{i0}\Big) \quad (i = 1, 2, \ldots, k-1) \qquad \text{and} \qquad p_k = 1\Big/\Big(1+\sum_{i=1}^{k-1}\lambda_{i0}\Big).$

Hence we have (Roy, 1953)

$$\sup_{\mathbf{p}'\neq\mathbf{p}_0'} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n}-n\mathbf{p}_0]}{\sqrt{\{\mathbf{a}'(\mathbf{p})\Lambda^0\mathbf{a}(\mathbf{p})\}}} = \sup_{\mathbf{b}'} \frac{\mathbf{b}'(\mathbf{p})[\mathbf{n}^*-n\mathbf{p}_0^*]}{\sqrt{\{\mathbf{b}'(\mathbf{p})\Lambda^0_{kk}\mathbf{b}(\mathbf{p})\}}} = +\{[\mathbf{n}^*-n\mathbf{p}_0^*]'\,\Lambda^{0-1}_{kk}[\mathbf{n}^*-n\mathbf{p}_0^*]\}^{\frac{1}{2}}.$$

We next observe that

$$\sigma^0_{ij} = -np_{i0}p_{j0} \quad \text{if} \quad i \neq j \qquad \text{and} \qquad \sigma^0_{ii} = np_{i0}(1-p_{i0}).$$

It is easy to check (Roy & Sarhan, 1956) that if $(\Lambda^0_{kk})^{-1} = (\alpha_{ij})$, then $\alpha_{ij} = 1/(np_{k0})$ if $i \neq j$ and $\alpha_{ii} = 1/(np_{i0}) + 1/(np_{k0})$. We have, therefore,

$$\sup_{\mathbf{p}'\neq\mathbf{p}_0'} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n}-n\mathbf{p}_0]}{\sqrt{\{\mathbf{a}'(\mathbf{p})\Lambda^0\mathbf{a}(\mathbf{p})\}}} = +\left[\sum_{i=1}^{k}\frac{(n_i-np_{i0})^2}{np_{i0}}\right]^{\frac{1}{2}}.$$

From (4·1·5) we see that (4·1·4) becomes

$$w(c) = \left\{\mathbf{n}' : +\left[\sum_{i=1}^{k}\frac{(n_i-np_{i0})^2}{np_{i0}}\right]^{\frac{1}{2}} \geqslant c\right\}. \tag{4·1·7}$$

Since the left-hand side of the inequality in (4·1·7) is essentially non-negative it is easy to see that we obtain a non-trivial solution only when $c > 0$. It is thus seen that the χ^2 critical region is obtained by using the union-intersection with respect to variation over the alternatives $\mathbf{p}'$ ($\neq \mathbf{p}_0'$), keeping fixed a quantity c defined (in terms of $\mathbf{p}_i'$ and n) by the right-hand side of (4·1·2). This means of course letting β vary with $\mathbf{p}$ in an appropriate manner. Now if n is large, from the asymptotic normality of the left-hand side of the inequality in (4·1·3), we have, as $n \to \infty$,

$$\beta(\mathbf{p}_1', c, n) \to \frac{1}{\sqrt{(2\pi)}}\int_c^\infty e^{-\frac{1}{2}t^2}\,dt. \tag{4·1·8}$$

In large samples it is thus seen that keeping c fixed means making β the same for all $\mathbf{p}_1$'s, which means that in large samples the χ^2 critical region (4·1·7) is a union-intersection critical region of type I (Roy, 1953). For large n_i's (the approximation would be good enough even for moderately large values of the n_i's) it is well known that the left-hand side of the inequality in (4·1·7) is asymptotically distributed as χ^2 with $(k-1)$ degrees of freedom. For a satisfactory proof see Cramér (1946, chapter 30).

4·2. *Test of a composite hypothesis against a composite alternative*

Suppose that the composite null hypothesis is given by

$$H_0: \quad \{p_i = p_i(\theta_1, \theta_2, \ldots, \theta_r)\}_{(\theta_1, \ldots, \theta_r) \in \Omega},$$

where $p_i(\theta_1, \ldots, \theta_r)$ are k known functions of r ($< k$) unknown parameters. The hypothesis does not specify the values of the parameters except that they belong to a certain parametric space Ω. The (composite) alternative is $H_1 \neq H_0$. For any given $(\theta_1^0, \theta_2^0, \ldots, \theta_r^0)$, we obtain, as in the previous section, a heuristic test of the hypothesis H_0': $\{p_i = p_i(\theta_1^0, \ldots, \theta_r^0)\}$ against $H_1' \neq H_0'$, which has a critical region

$$w(c, \theta_1^0, \ldots, \theta_r^0) = \left\{\mathbf{n}: \sum_{i=1}^{k} \frac{[n_i - np_i(\theta_1^0, \ldots, \theta_r^0)]^2}{np_i(\theta_1^0, \ldots, \theta_r^0)} \geqslant c^2\right\}. \tag{4·2·1}$$

This critical region is the region of rejection of H_0': $\{p_i = p_i(\theta_1^0, \ldots, \theta_r^0)\}$ for a specific $(\theta_1^0, \ldots, \theta_r^0)$. Now to reject H_0: $\{p_i = p_i(\theta_1, \ldots, \theta_r)\}_{(\theta_1, \ldots, \theta_r) \in \Omega}$ would be to reject

$$H_0'\text{: } p_i = p_i(\theta_1^0, \ldots, \theta_r^0) \text{ for every } (\theta_1^0, \ldots, \theta_r^0) \in \Omega,$$

and thus using the union-intersection principle for the second time we have for H_0 the critical region

$$\begin{aligned} w(c, H_0) &= \bigcap_{(\theta_1, \ldots, \theta_r) \in \Omega} w(c, \theta_1, \ldots, \theta_r) \\ &= \left\{\mathbf{n}: \inf_{(\theta_1, \ldots, \theta_r) \in \Omega} \sum_{i=1}^{k} \frac{[n_i - np_i(\theta_1, \ldots, \theta_r)]^2}{np_i(\theta_1, \ldots, \theta_r)} > c^2\right\}, \end{aligned} \tag{4·2·2}$$

which is precisely the minimum χ^2 critical region. The equations giving the θ_j's in terms of the n_i's, in the form for minimum χ^2 are

$$\frac{\partial}{\partial \theta_j} \sum_{i=1}^{k} \frac{[n_i - np_i(\theta_1, \ldots, \theta_r)]^2}{np_i(\theta_1, \ldots, \theta_r)} = 0 \quad (j = 1, 2, \ldots, r). \tag{4·2·3}$$

It has been shown by Cramér that for large n_i's the equations (4·2·3) can be replaced by the maximum-likelihood equations which are simpler to use, the likelihood function being

$$\phi_0 = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}(\theta_1, \ldots, \theta_r). \tag{4·2·4}$$

The maximum-likelihood equations can be put in the form

$$0 = \frac{\partial}{\partial \theta_j} \log \phi_0 = \sum_{i=1}^{k} \frac{n_i}{np_i} \frac{\partial p_i}{\partial \theta_j} = \sum_{i=1}^{k} \frac{n_i - np_i}{np_i} \frac{\partial p_i}{\partial \theta_j} \quad (j = 1, 2, \ldots, r). \tag{4·2·5}$$

5. Large-sample tests of the hypotheses in §2

Assuming that $\sum_i n_{ij} = n_{0j}$ (fixed) and $\sum_i p_{ij} = p_{0j} = 1$ (where i itself may be a multiple subscript like $i_1 i_2 \ldots i_k$ and j may be a multiple subscript like $j_1 j_2 \ldots j_l$) and using the traditional approach of conditional probability, we have, under any hypothesis which expresses the p_{ij}'s in terms of a lesser number of free or nuisance parameters, the result that, as $n \to \infty$ subject to n_{0j}/n being held constant, $\sum_{i,j} (n_{ij} - n_{0j}\hat{p}_{ij})^2/(n_{0j}\hat{p}_{ij})$ tends to have the

χ^2-distribution with degrees of freedom equal to the numbers of cells minus the number of linear constraints on the n_{ij}'s (which we shall here replace by the number of separate multinomial distributions) minus the number of independent parameters that are estimated from the data. Notice that the $\hat{p}_{ij}$'s are the maximum-likelihood estimates of the p_{ij}'s obtained by expressing the p_{ij}'s in terms of the lesser number of free parameters from the hypothesis, then substituting in ϕ, and then maximizing the ϕ with respect to these free or nuisance parameters. However, here we do not rely on the traditional approach but upon a proof of the above theorem which starts from

$$\phi = \prod_j \left[\frac{n_{0j}!}{\prod_i n_{ij}!} \prod_i p_{ij}^{n_{ij}} \right],$$

and proceeds along the lines followed by Cramér. This, among other theorems, is given in a paper to be shortly submitted to *Biometrika*.

5·1. *The problem of* § 2·1

We consider § 2·1, start from (2·1·1), maximize $\log \phi_0$ with respect to the p_{i0}'s and p_{0j}'s subject to $\sum_i p_{i0} = \sum_j p_{0j} = 1$ (using Lagrangian multipliers) and obtain the maximum-likelihood solutions: $\hat{p}_{i0} = n_{i0}/n$ and $\hat{p}_{0j} = n_{0j}/n$. The number of independent parameters estimated from the data is $r+s-2$, and hence by § 5, the test of independence here is based on a statistic which has the χ^2-distribution with degrees of freedom

$$rs - 1 - (r+s-2) = (r-1)(s-1)$$

and whose form is

$$\sum_{i,j} \frac{\left(n_{ij} - n\frac{n_{i0}}{n}\frac{n_{0j}}{n}\right)^2}{n\frac{n_{i0}}{n}\frac{n_{0j}}{n}} = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_{i0}n_{0j}}{n}\right)^2}{\frac{n_{i0}n_{0j}}{n}}. \qquad (5\cdot1\cdot1)$$

5·2. *The problem of* § 2·2

We start from (2·2·1) and maximize $\log \phi_0$ with respect to q_{0j}'s subject to $\sum_j q_{0j} = 1$ and obtain the maximum-likelihood solutions: $\hat{q}_{0j} = n_{0j}/n$. The number of independent parameters estimated from the data is $s-1$, and hence by § 5 the test here is to be based on a statistic which has the χ^2-distribution with degrees of freedom

$$r(s-1) - (s-1) = (r-1)(s-1)$$

and whose form is

$$\sum_i \sum_j \frac{\left(n_{ij} - n_{i0}\frac{n_{0j}}{n}\right)^2}{n_{i0}\frac{n_{0j}}{n}} = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_{i0}n_{0j}}{n}\right)^2}{\frac{n_{i0}n_{0j}}{n}}. \qquad (5\cdot2\cdot1)$$

5·3. *The problem of* § 2·3

We start from (2·3). Here we have already, under the null hypothesis, $p_{ij} = n_{i0}n_{0j}/n^2$, and using the remarks of § 5 we note that the test is based on a statistic having the χ^2-distribution with degrees of freedom $rs-(r+s-1) = (r-1)(s-1)$ and the form

$$\sum_{i,j} \left(n_{ij} - \frac{n_{i0}n_{0j}}{n}\right)^2 \Big/ \frac{n_{i0}n_{0j}}{n}. \qquad (5\cdot3\cdot1)$$

6. Large-sample tests of the hypotheses in §3

6·1. *The problems of* § 3·1, *i.e. where* '*i*', '*j*' *and* '*k*' *are all* '*variates*'

6·1*a*. *The problem of* § 3·1*a*

Independence between '*i*' *and* '*j*' *for fixed* '*k*'. Under H_0 of (3·1·1) we shall have

$$\phi_0 \propto \prod_{i,j,k} (p_{i0k}p_{0jk}/p_{00k})^{n_{ijk}}. \tag{6·1·1}$$

To test the hypothesis here we maximize $\log \phi_0$ with respect to the p_{i0k}'s, p_{0jk}'s and p_{00k}'s subject to $\sum_i p_{i0k} = \sum_j p_{0jk} = p_{00k}$ and $\sum_k p_{00k} = 1$, and obtain the maximum-likelihood solutions $\hat{p}_{i0k} = n_{i0k}/n$, $\hat{p}_{0jk} = n_{0jk}/n$ and $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from this data is $(r-1)t+(s-1)t+(t-1)$. And hence by § 5 the test of conditional independence is here based on a statistic which has the χ^2-distribution with degrees of freedom $rst-1-t(r-1)-t(s-1)-(t-1) = t(r-1)(s-1)$ and whose form is

$$\sum_{i,j,k} \left(n_{ijk} - \frac{n_{i0k}n_{0jk}}{n_{00k}}\right)^2 \Big/ \frac{n_{i0k}n_{0jk}}{n_{00k}}. \tag{6·1·2}$$

Independence between '*i*' *and* '*k*' *and also between* '*j*' *and* '*k*'. This can be handled exactly on the lines of § 5 and will not be discussed separately.

Independence between '*i*', '*j*' *and* '*k*'. To test this we start from the hypothesis of (3·1·3) giving

$$\phi_0 \propto \prod_{i,j,k} (p_{i00}p_{0j0}p_{00k})^{n_{ijk}}, \tag{6·1·3}$$

maximize $\log \phi_0$ with respect to p_{i00}'s, p_{0j0}'s and p_{00k}'s subject to $\sum_i p_{i00} = \sum_j p_{0j0} = \sum_k p_{00k} = 1$, and obtain the maximum-likelihood solutions: $\hat{p}_{i00} = n_{i00}/n$, $\hat{p}_{0j0} = n_{0j0}/n$ and $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from the data is $(r+s+t-3)$, and hence by § 5 the test is here based on a statistic which has the χ^2-distribution with degrees of freedom $rst-1-(r+s+t-3) = rst-r-s-t+2$, and whose form is

$$\sum_{i,j,k} \left(n_{ijk} - \frac{n_{i00}n_{0j0}n_{00k}}{n^2}\right)^2 \Big/ \frac{n_{i00}n_{0j0}n_{00k}}{n^2}. \tag{6·1·4}$$

6·1*b*. *The problems of* § 3·1*b*

Independence between '(i,j)' *and* '*k*'. Under H_0 of (3·1·4) we shall have

$$\phi_0 \propto \prod_{i,j,k} (p_{ij0}p_{00k})^{n_{ijk}}. \tag{6·1·5}$$

To test this hypothesis we maximize $\log \phi_0$ with respect to p_{ij0}'s and p_{00k}'s subject to $\sum_{i,j} p_{ij0} = \sum_k p_{00k} = 1$ and obtain the maximum-likelihood solutions: $\hat{p}_{ij0} = n_{ij0}/n$ and $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from the data is $(rs-1)+(t-1)$ and hence by § 5 the test is based on a statistic having the χ^2-distribution with degrees of freedom $rst-1-[(rs-1)+(t-1)] = (rs-1)(t-1)$ and having the form

$$\sum_{i,j,k} \left(n_{ijk} - \frac{n_{ij0}n_{00k}}{n}\right)^2 \Big/ \frac{n_{ij0}n_{00k}}{n}. \tag{6·1·6}$$

Independence between 'i' and 'k' and between 'j' and 'k'. Since this can be handled on the same lines as in § 5, it will not be separately discussed.

The 'no interaction' hypothesis of (3·1·6). This has been discussed in detail in another paper by Roy & Kastenbaum (1956) and will not be discussed here. The test will be based on a statistic having the χ^2-distribution with degrees of freedom $(r-1)(s-1)(t-1)$ and having a rather complicated form (see Bartlett, 1935; Norton, 1945; Roy & Kastenbaum, 1956).

6·2. *The problems of* § 3·2, *i.e. when 'i' and 'j' are 'variates' and 'k' is a 'way of classification'*

6·2*a*. *The problems of* § 3·2*a*

Independence between 'i' and 'j' for each k. Under H_0 of (3·2·1) we start from

$$\phi_0 \propto \prod_{i,j,k} (p_{i0k} p_{0jk})^{n_{ijk}} \tag{6·2·1}$$

and maximize $\log \phi_0$ with respect to the p_{i0k}'s and p_{0jk}'s subject to $\sum_i p_{i0k} = \sum_j p_{0jk} = p_{00k} = 1$, and obtain the maximum-likelihood solutions: $\hat{p}_{i0k} = n_{i0k}/n_{00k}$, $\hat{p}_{0jk} = n_{0jk}/n_{00k}$. The number of independent parameters to be estimated from these data is $t(r-1)+t(s-1)$, and hence the test here is to be based on a statistic having the χ^2-distribution with degrees of freedom $t(rs-1)-t(r-1)-t(s-1) = t(r-1)(s-1)$ and having the form

$$\sum_k \left[\sum_{i,j} \left(n_{ijk} - n_{00k} \frac{n_{i0k} n_{0jk}}{n_{00k}^2} \right)^2 \Big/ \left(n_{00k} \frac{n_{i0k} n_{0jk}}{n_{00k}^2} \right) \right]. \tag{6·2·2}$$

The problems under (3·2·2) or (3·2·3) will not be discussed separately.

6·2*b*. *The problems of* § 3·2*b*

The hypothesis that p_{ijk} is independent of 'k'. Under H_0 of (3·2·4) we start from

$$\phi_0 \propto \prod_{i,j,k} q_{ij0}^{n_{ijk}}, \tag{6·2·3}$$

maximize $\log \phi_0$ with respect to the q_{ij0}'s subject to $\sum_{i,j} q_{ij0} = 1$, and obtain the maximum-likelihood solutions: $\hat{q}_{ij0} = n_{ij0}/n$. The number of independent parameters to be estimated from the data is $(rs-1)$, and hence by § 5 the test is to be based on a statistic having the χ^2-distribution with degrees of freedom $t(rs-1)-(rs-1) = (rs-1)(t-1)$ and having the form

$$\sum_k \left[\sum_{i,j} \left(n_{ijk} - n_{00k} \frac{n_{ij0}}{n} \right)^2 \Big/ \left(n_{00k} \frac{n_{ij0}}{n} \right) \right]. \tag{6·2·4}$$

The problems under (3·2·5) or (3·2·6) will not be separately discussed here.

6·3. *The problems of* § 3·3, *i.e. when 'i' is a 'variate' and 'j' and 'k' are 'ways of classification'*

6·3*a*. *The problem of* § 3·3*a*

The hypothesis that for any 'k', p_{ijk} is independent of 'j'. Under H_0 of (3·3·1) we start from

$$\phi_0 \propto \prod_{i,j,k} q_{i0k}^{n_{ijk}} \tag{6·3·1}$$

and maximize $\log \phi_0$ with respect to the q_{i0k}'s subject to $\sum_i q_{i0k} = 1$, and obtain the maximum-likelihood solutions $\hat{q}_{i0k} = n_{i0k}/n_{00k}$. The number of independent parameters to be estimated

from the data is $t(r-1)$, and hence by § 5 the test is to be based on a statistic having the χ^2-distribution with degrees of freedom $st(r-1)-t(r-1)=t(r-1)(s-1)$ and having the form

$$\sum_{i,k}\left[\sum_{i}\left(n_{ijk}-n_{0jk}\frac{n_{i0k}}{n_{00k}}\right)^2\Big/\left(n_{0jk}\frac{n_{i0k}}{n_{00k}}\right)\right]. \tag{6·3·2}$$

6·3*b*. *The problem of* § 3·3*b*

This will be exactly on the lines as the previous case and will not be discussed separately. We shall also omit a discussion of the problem under (3·3·3).

6·4. *The problems of* § 3·4, *i.e. when* '*i*' *is a variate and* '*j*' *and* '*k*' *are ways of classification in the sense of an incomplete design*

The hypothesis that p_{ijk} *is independent of* '*j*'. We start from (3·4), put $p_{ijk}=q_{i0k}$ and thus have $\phi_0 \propto \prod_{i,j,k} q_{i0k}^{n_{ijk}}$. On maximizing $\log\phi_0$ with respect to the q_{i0k}'s subject to $\sum_i q_{i0k}=1$, we obtain a solution for the q_{i0k}'s in terms of the n_{ijk}'s which is a rather complicated set of functions of n_{ijk}'s of the same structure as the corresponding least-squares solutions in linear estimation. One or two such solutions for some linked block designs will be discussed in a later paper. However, this solution, inserted in the χ^2 functions, will have the χ^2-distribution with degrees of freedom $(r-1)\sum_{j=1}^{s} t_j-(rt-1)$.

The hypothesis that p_{ijk} *is independent of* '*j*', can be handled on exactly similar lines and need not be separately considered.

7. Concluding remarks

The extension from three to more dimensions does not present any essentially new problems. One or two additional features of interest of such extensions will be discussed later.

The authors believe that this is an attempt at a somewhat systematic exposition (i) which is based on a clear distinction between a 'variate' and a 'way of classification', that stems from differing experimental situations and sampling schemes, (ii) which sets up different probability models for the different situations, and (iii) which poses different types of hypothesis according as it is a 'multivariate analysis' situation or an 'analysis of variance' situation or something of a mixed type. Every type of experimental situation has its own appropriate sampling scheme and appropriate probability model, and the authors believe that it would not be proper to force some kind of sampling scheme and probability model on the wrong kind of experimental situation and also to pose one kind of hypothesis on the wrong kind of sampling scheme and probability model.

For two analogous null hypotheses under two different probability models we have eventually the same χ^2 with the same distribution under the respective null hypotheses. But the powers of these tests, that is, the distributions of these statistics under the respective non-null hypotheses, are entirely different. In large samples these powers would of course, in the usual sense, tend to 1 in each case. But the asymptotic powers (in the sense of Pitman and Lehmann, which is perhaps the only sense that is relevant here) can be obtained and compared. They have been obtained; they are different and will be given a later paper.

In many situations in which a particular hypothesis H_0 with an associated χ^2 is the intersection of several hypothesis $H_{01}, H_{02}, \ldots$, with associated $\chi^2_1, \chi^2_2, \ldots$, it so happens that

$\chi^2 = \chi_1^2 + \chi_2^2 + \ldots$ and $\chi_1^2, \chi_2^2, \ldots$ are also independently distributed, but the additivity is not in the usual algebraic sense; it is only in probability and asymptotically as $n \to \infty$, and the independence is also in the asymptotic sense. Take, for example, the hypotheses (3·1·1), (3·1·2) and (3·1·3) and let us call them H_{01}, H_{02} and H_{03}, and H_0. We note that $H_0 = H_{01} \cap H_{02} \cap H_{03}$. Let the associated χ^2's be denoted by $\chi_1^2, \chi_2^2, \chi_3^2$ and χ^2. Then, in this case it has been shown that, in large samples and under the null hypothesis, χ_1^2, χ_2^2 and χ_3^2 are independently distributed and $\chi_1^2 + \chi_2^2 + \chi_3^2 \to \chi^2$ in probability. We have an exactly similar situation for the group of hypotheses (3·1·4), (3·1·5) and (3·1·6). These are situations in multivariate analysis. There are similar situations in analysis of variance also, for example, with the group of hypotheses (3·2·4), (3·2·5) and (3·2·6) or with the group (3·3·1), (3·3·2) and (3·3·3). But this will not be true, for example, with a similar group of hypotheses on an incomplete block design or general types of designs indicated in §3·4. A proof of 'independence' and 'additivity' covering the problems of this paper and some other problems (where 'independence' and 'additivity' exist) will be given in a later paper.

We have defined the hypothesis of 'no interaction' between 'i' and 'j' when 'i', 'j' and 'k' are all 'variates' or 'i' and 'j' are variates and 'k' is a way of classification. It is a kind of bridge over the gap between (*a*) the hypotheses of independence of 'i' and 'k', 'j' and 'k', and (*b*) the hypotheses of independence of '(i,j)' and 'k'. Through this mechanism we get a formula of which the special cases for $2 \times 2 \times 2$ and $2 \times 2 \times t$ were obtained by earlier workers (Bartlett, 1935; Norton, 1945). They obtained the formula through presumably some different mechanism, and perhaps for the analysis of variance situations where 'i' and 'j' are 'ways of classification' and 'k' is a 'variate'. The authors feel that for the analysis of variance situations with frequency types of data the hypothesis of 'no interaction' in the form in which it is usually posed and tested may not be too meaningful, and they doubt whether, even with 'normal variate' type of data, it is very useful. The controversy that has raged for many years now over questions of interpretation is not without its own lessons. The authors feel this way about the whole concept of interaction in analysis of variance. Looking for a possible motivation behind the customary (and mostly 'normal variate') analysis, one cannot help feeling that factorial experiments (whether on the 'normal variate' type of data or more general types of data) present a problem which is essentially different from that of the rest of analysis of variance, e.g. the usual tests of significance of treatment differences. Assume, for simplicity, in the beginning that there is just one factor at, say, k levels. One might regard these as treatments, and test whether there are significant differences between these, or, in terms of some characteristic, pick out the 'best' among these or rank these in some order. But that does not appear to be the relevant question here. The (second) characteristic in terms of which we have the k levels seems to be a continuous variate which is observed at k levels for practical convenience and what is of interest seems to be to lay down a statistical rule by which we can decide about the 'best' or 'optimum' point in the range of the continuous variate, the decision rule being in terms of observations at discrete levels and the 'best' or 'optimum' being in relation to the first characteristic. Likewise, taking, for example, two factors at k and l levels respectively, the problem seems to be not to test whether there are significant differences between these kl combinations regarded as treatments (which would really be a linear problem) or to pick out the 'best' among these or to rank these (in terms of some characteristic), which again would be each a really linear problem. It seems that there is a (second) characteristic in terms of which we have the k levels of the first factor and a third characteristic in terms of which we have

the l levels of the second factor, both these (second and third) characteristics being supposed to be continuous variates. The problem is to lay down a statistical rule by which we can decide about the 'best' or 'optimum' point (in relation to the first characteristic) on the plane of the second and third characteristics, regarded as two continuous variates, the decision rule being in terms of observations at the kl discrete level combinations. This of course can be generalized to several factors. The customary analysis into 'main effects', 'interactions' of various orders, confounding, etc., all seem to point very stongly in this direction. Some work has already been done from this standpoint and further work is under way. This will be discused in a later monograph.

In conclusion, it is a great pleasure to thank the editor and the referees for their valuable suggestions for the improvement of the paper both in form and in content.

REFERENCES

BARNARD, G. A. (1947). Significance tests for 2×2 tables. *Biometrika*, **34**, 123–38.

BARTLETT, M. S. (1935). Contingency table interactions. *J. R. Statist. Soc. Suppl.* **2**, 248–52.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

FISHER, R. A. (1922). On the interpretation of χ^2 from contingency tables and the calculation of P. *J. R. Statist. Soc.* **85**, 87–94.

NEYMAN, J. (1949). Contribution to the theory of the χ^2 test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 239–73. University of California Press.

NORTON, H. W. (1945). Calculation of chi-square for complex contingency tables. *J. Amer. Statist. Ass.* **40**, 251–8.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139–67.

REIERSØL, O. (1954). Tests of linear hypotheses concerning binomial experiments. *Skand. Aktuar-Tidskr.* **37**, 38–59.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–38.

ROY, S. N. & KASTENBAUM, M. A. (1956). On the hypothesis of no interaction in a multiway contingency table. *Ann. Math. Statist.* **27**, 749–57

ROY, S. N. & SARHAN, A. E. (1956). On inverting a class of patterned matrices. *Biometrika*, **43**, 227–31.

YULE, G. U. & KENDALL, M. G. (1950). *An Introduction to the Theory of Statistics*. 14th ed. London: Chas. Griffin and Co.

A TWO-SAMPLE DISTRIBUTION-FREE TEST

By A. R. KAMAT

Fergusson College, University of Poona

1. In 1947 Mann & Whitney proposed a test for the equivalence of two distribution functions which was based on the sample ranks. Given two samples of size m and n they proposed that the two samples be pooled and ranked in ascending (or descending) order of magnitude. The sum of the ranks of one of the samples is then used in a test of the hypothesis that the two samples have come from the same population or, if it is preferred, of the hypothesis of the equivalence of the two distribution functions. Such a criterion will clearly be most sensitive to possible differences of location of the two distribution functions. Rosenbaum (1953) assumed that the location parameters of the distribution functions were equal, and the criterion he suggested could be used to test for the equivalence of parameters of dispersion, by counting the numbers in one sample falling outside the range of the other. An alternative test to Rosenbaum's is discused here, using a criterion based on the rank ranges of both samples.

2. It is assumed that there are two samples $\{x_i\}$ $(i = 1, 2, \ldots, n)$ and $\{y_j\}$ $(j = 1, 2, \ldots, m;$ $m \geqslant n)$. The location parameters of the distributions of x and y are assumed to be equal. The samples are pooled and arranged in order of magnitude. Let R_n and R_m be the range of ranks of x and y respectively. The test statistic proposed is

$$D_{n,m} = R_n - R_m + m, \tag{1}$$

where $D_{n,m}$ can take values $0, 1, \ldots, m+n$. Large or small values of $D_{n,m}$ will indicate possible divergence from the hypothesis that the parameters of dispersion of the populations from which the samples have been drawn are equal.

3. The following example illustrates the calculation of $D_{n,m}$. Two analysts in the same laboratory made repeated determinations of the percentage fibre in a sample of soya cattle cake which had been ground down into a uniformly mixed powder. The first analyst made $n = 8$ determinations and the second $m = 10$. The observations were as follows, the ranks of each, from 1 to 18, after pooling being shown.* It will be seen that

$$D_{8,10} = 12 - 17 + 10 = 5.$$

The question is whether this result suggests a significant difference in the accuracy with which the two men can repeat their analyses:

Analyst 1:	Observations	12·38	12·53	12·25	12·37	12·48	12·43	12·30	12·46	
	Pooled rank	8	2	14	9	3	6	12	4	
Analyst 2:	Observations	12·45	12·31	12·31	12·20	12·26	12·73	12·42	12·17	12·09	12·23
	Pooled rank	5	10·5	10·5	16	13	1	7	17	18	15

$n = 8$, $R_n = 14 - 2 = 12$; $m = 10$, $R_m = 18 - 1 = 17$.

If we enter the table of percentage points (p. 378 below) with $m + n = 18$, $n = 8$, it is seen that for a two-tailed test, using a 5 % significance level, the lower and upper limits for $D_{8,10}$ are 3 and 15 respectively, so that the observed value of 5 is not significant.

* There is one tie, i.e. two readings of 12·31 % by analyst 2, but this does not affect the value of $D_{8,10}$.

Table 1. *Percentage points of* $D_{n,m}$

For the lower points, $P\{D_{n,m} \leqslant D_\alpha\} \leqslant \alpha$; for the upper points $P\{D_{n,m} \geqslant D_\alpha\} \leqslant \alpha$, with $100\alpha = 0{\cdot}5$, $1{\cdot}0$, $2{\cdot}5$, $5{\cdot}0$. A stroke indicates that no value of D_α satisfies the inequality.

$m+n$	n	Lower percentage points				Upper percentage points			
		0·5	1·0	2·5	5·0	5·0	2·5	1·0	0·5
7	2	—	—	—	—	6	—	—	—
	3	—	—	—	—	—	—	—	—
8	2	—	—	—	—	7	—	—	—
	3	—	—	—	—	8	—	—	—
	4	—	—	—	0	8	—	—	—
9	2	—	—	—	—	8	—	—	—
	3	—	—	—	—	9	9	—	—
	4	—	—	—	0	9	9	—	—
10	2	—	—	—	—	9	9	—	—
	3	—	—	—	—	10	10	—	—
	4	—	—	0	0	10	10	—	—
	5	—	—	0	0	10	10	—	—
11	2	—	—	—	—	10	10	—	—
	3	—	—	—	0	11	11	—	—
	4	—	—	0	0	10	11	11	—
	5	—	—	0	1	10	11	11	—
12	2	—	—	—	—	10	11	—	—
	3	—	—	—	0	11	12	12	—
	4	—	—	0	0	11	12	12	—
	5	—	0	0	1	11	12	12	—
	6	—	0	0	1	11	12	12	—
13	2	—	—	—	—	11	12	—	—
	3	—	—	—	0	11	13	13	—
	4	—	—	0	1	12	13	13	13
	5	—	0	1	1	12	12	13	13
	6	0	0	1	1	12	12	13	13
14	2	—	—	—	—	12	13	—	—
	3	—	—	—	0	12	14	14	—
	4	—	0	0	1	13	13	14	14
	5	0	0	1	2	13	13	14	14
	6	0	0	1	2	12	13	14	14
	7	0	0	1	2	12	13	14	14
15	2			—	—	12	14	14	—
	3	—	—	0	0	13	15	15	15
	4	—	0	0	1	14	14	15	15
	5	0	0	1	2	13	14	15	15
	6	0	1	1	2	13	14	14	15
	7	0	1	2	2	13	13	14	15
16	2	—	—	—	—	13	14	15	—
	3	—	—	0	0	14	15	16	16
	4	—	0	1	1	15	15	16	16
	5	0	0	1	2	14	15	16	16
	6	0	1	2	3	14	14	15	16
	7	1	1	2	3	13	14	15	15
	8	1	1	2	3	13	14	15	15
17	2	—	—	—	—	14	15	16	—
	3	—	—	0	0	15	15	17	17
	4	—	0	1	2	15	16	17	17
	5	0	1	2	2	15	16	16	17
	6	1	1	2	3	15	15	16	16
	7	1	1	2	3	14	15	16	16
	8	1	2	3	3	14	15	15	16
18	2	—	—	—	—	15	16	17	—
	3	—	—	0	1	16	16	18	18
	4	0	0	1	2	16	17	17	18
	5	0	1	2	3	16	16	17	17
	6	1	1	2	3	15	16	17	17
	7	1	2	3	4	15	16	16	17
	8	1	2	3	4	15	15	16	17
	9	1	2	3	4	14	15	16	17
19	2	—	—	—	—	15	17	18	—
	3	—	—	0	1	16	17	19	19
	4	0	0	1	2	16	18	18	19
	5	0	1	2	3	17	17	18	18
	6	1	2	3	4	16	17	18	18
	7	1	2	3	4	16	16	17	18
	8	2	2	3	4	15	16	17	17
	9	2	2	3	4	15	16	17	17
20	2	—	—	—	—	16	18	19	—
	3	—	—	0	1	17	18	20	20
	4	0	0	1	2	17	19	19	20
	5	1	1	2	3	18	18	19	19
	6	1	2	3	4	17	18	18	19
	7	2	2	4	5	17	17	18	18
	8	2	3	4	5	16	17	18	18
	9	2	3	4	5	16	17	17	18
	10	2	3	4	4	16	16	17	18

4. The probability distribution of $D_{n,m}$ follows from the application of simple combinatorial rules. The total number of ways in which the sequence can be arranged is $\binom{m+n}{n}$ and this will be therefore the totality of the values of $D_{n,m}$. This totality of values can also be built up as follows:

(*a*) $R_n = n-1$, $R_m = m-1$; $D_{n,m} = n$.

There are only two ways in which this can happen. Either the x-sample takes the first n places or the last n.

(*b*) $R_n = n-1+i$ $(i = 0, 1, \ldots, m-1)$, $R_m = m-1+n$; $D_{n,m} = i$.

Let the y-sample occupy the first and the last places and the x-sample n of the $m+n-2$ remaining places, remembering that R_n must equal $n-1+i$. This can be done in

$$(m-1-i)\binom{n+i-2}{n-2} \text{ ways.}$$

(See, for instance, Whitworth (1901), *Choice and Chance*, chapter III.)

(*c*) $R_n = n-1+m$, $R_m = m-1+j$, $(j = 0, 1, \ldots, n-1)$; $D_{n,m} = n+m-j$.

By symmetry the number of arrangements is

$$(n-1-j)\binom{m+j-2}{m-2} \text{ ways.}$$

(*d*) $R_n = n-1+i$, $R_m = m-1+j$; $D_{n,m} = n+i-j$.

Excluding (*a*), (*b*) and (*c*) it can be shown that the number of ways will be

$$2\binom{i+j-2}{j-1} \quad (i = 1, 2, \ldots, (m-1); j = 1, 2, \ldots, (n-1)).$$

Combining these four cases we may write

$$P\{D_{n,m} = n+i-j = r\} = \frac{1}{\binom{m+n}{n}}\left\{2A_r \sum_j \binom{r-n+2j-2}{r-n+j-1} + 2B_r \sum_i \binom{n-r+2i-2}{n-r+i-1}\right.$$

$$\left. + C_r(m-1-r)\binom{n+r-2}{n-2} + D_r(r-m-1)\binom{2m+n-r-2}{m-2} + 2E_r\right\}, \quad (2)$$

where $A_r = 1$ for $r \leqslant m$, $B_r = 1$ for $r > m$, $C_r = 1$ for $r < m$, $D_r = 1$ for $r > m$, $E_r = 1$ for $r = n$ and they are zero otherwise. When the two samples are of the same size a certain simplification of the probability distribution function is possible. Write $m = n$, whence

$$P\{D_{n,n} = r\} = \frac{1}{\binom{2n}{n}}\left\{2\sum_j \binom{r-n+2j-2}{r-n+j-1} + C_r(n-1-r)\binom{n+r-2}{n-2} + 2E_r\right\} \quad (3)$$

for $r = 0, 1, \ldots, n$. $C_r = 1$ for $r < n$, $E_r = 1$ for $r = n$ and they are both zero otherwise. Values of the percentage points for $m+n \leqslant 20$ are given in Table 1. They are calculated from the distribution function of $D_{n,m}$ given above.

The calculation of percentage points from the exact distribution clearly becomes impracticable when the sequence becomes at all large.

5. The moments of the distribution of $D_{n,m}$ can be obtained from the distribution itself or by considering the moments of R_m and R_n and their cross-products. Whichever way is chosen, the combinatorial sums are easily reduced by making use of the relation

$$\sum_{i=0}^{s}\binom{k+s}{k}=\binom{k+s+1}{k+1},$$

although the reduction of the resulting polynomials is a long process. The first three moments of $D_{n,m}$ are

$$\left.\begin{aligned}
\mu_1' &= m-\frac{2m}{n+1}+\frac{2n}{m+1},\\
\mu_2 &= \frac{2m(n-1)(m+n+1)}{(n+1)^2(n+2)}+\frac{2n(m-1)(m+n+1)}{(m+1)^2(m+2)}+\frac{8mn}{(m+1)(n+1)}-4+4\Big/\binom{m+n}{n},\\
\mu_3 &= -\frac{2m(n-1)(n-3)(m+n+1)(2m+n+1)}{(n+1)^3(n+2)(n+3)}+\frac{2n(m-1)(m-3)(m+n+1)(m+2n+1)}{(m+1)^3(m+2)(m+3)}\\
&\quad+\frac{24(m-n)(m-1)(n-1)\{(m+n+1)(m+n+2)-mn\}}{(m+1)^2(n+1)^2(m+2)(n+2)}\\
&\quad+\frac{12(m-n)}{(m+1)(n+1)}\left\{\frac{4mn}{(m+1)(n+1)}-2-[(m-1)(n-1)-2]\Big/\binom{m+n}{n}\right\}.
\end{aligned}\right\} \tag{4}$$

The fourth moment was also worked out but is not reproduced here. For the symmetrical case

$$\left.\begin{aligned}
\mu_1' &= n,\\
\mu_2 &= \frac{4n(4n-1)}{(n+1)(n+2)}-4+4\Big/\binom{2n}{n},\\
\mu_3 &= 0.
\end{aligned}\right\} \tag{5}$$

6. Figs. 1 and 2 show the form of the distribution of $D_{n,m}$ for $m+n=16$, $m=8$ and 12. The distribution is trimodal and it is not expected that the beta ratios, some values for which are given in Tables 2 and 3, will be very helpful in deriving approximations by which to extend the percentage limits of Table 1 beyond $m+n=20$.* Some light is, however, thrown on the problem by considering the limiting distribution of $D_{n,m}$.

7. The limiting distribution has been obtained by D. E. Barton in an addendum p. 386 following this paper. Let $m+n=N$ and let N increase so that

$$m/N\to p, \quad n/N\to q=1-p.$$

If we write $d=D_{n,m}-m$, the limiting distribution of d is

$$\left.\begin{aligned}
f(d) &= p^{-d}q^2\{2p^2/(1-Q)-1-d\} \quad (d<0)\\
&= 2Q^2/(1-Q) \quad (d=0)\\
&= q^dp^2\{2q^2/(1-Q)-1+d\} \quad (d>0),
\end{aligned}\right\} \tag{6}$$

* Use of the Pearson–Merrington table of standardized deviates (Pearson & Hartley, 1954, Table 42) with the correct mean, standard deviation and beta values and the application of a continuity correction, has been found to give surprisingly accurate estimates of the 2·5% points at $m+n=16$. The possibilities of this method of approximation, however, need further investigation.

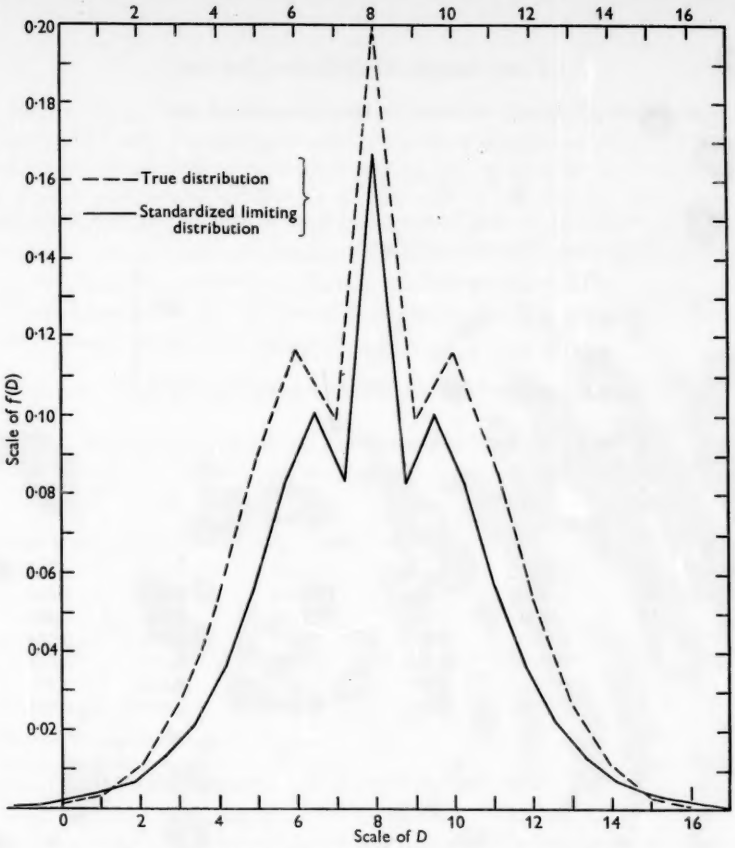


Fig. 1. Distribution of $D_{n,m}$ for case $m+n=16$, $m=8$.

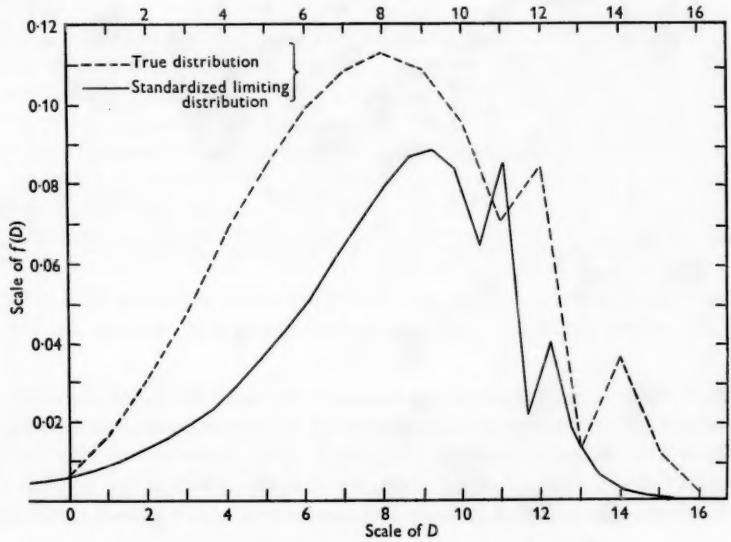


Fig. 2. Distribution of $D_{n,m}$ for case $m+n=16$, $m=12$.

where $Q = pq$. When $p = q = \frac{1}{2}$, we have for the symmetrical case

$$\left.\begin{aligned} f(d) &= \tfrac{1}{6} && (d=0) \\ &= \tfrac{1}{12}\, 2^{-|d|}\,(3\,|\,d\,|-1) && (d \neq 0). \end{aligned}\right\} \tag{7}$$

The first four moments of the limiting distribution can be derived from the values obtained in the finite case. They are as follows:

$$\left.\begin{aligned} \mathscr{E}(d) &= 2(q-p)\,Q^{-1} \\ \mu_2(d) &= 4 - 6Q^{-1} + 2Q^{-2}, \\ \mu_3(d) &= 2(q-p)\,\{Q^{-1} - 5Q^{-2} + 2Q^{-3}\}, \\ \mu_4(d) &= 28 - 174Q^{-1} + 266Q^{-2} - 144Q^{-3} + 24Q^{-4}. \end{aligned}\right\} \tag{8}$$

Table 2. *Momental constants of $D_{n,m}$ for case $m+n = 16$*

n	m	Mean	S.D.	μ_2	$\sqrt{\beta_1}$	β_2
2	14	4·9333	3·805	14·4789	0·669	2·681
3	13	6·9250	3·616	13·0765	0·178	2·340
4	12	7·8154	3·271	10·7022	0·000	2·404
5	11	8·1667	2·993	8·9567	−0·069	2·545
6	10	8·2338	2·799	7·8355	−0·072	2·599
7	9	8·1500	2·687	7·2188	−0·044	2·615
8	8	8·0000	2·650	7·0225	0·000	2·618

Table 3. *Momental constants of $D_{n,m}$ (symmetrical case)*

$n=m$	S.D.	μ_2	β_2
2	1·155	1·3333	3·000
3	1·673	2·8000	2·500
4	2·014	4·0571	2·413
5	2·250	5·0635	2·434
7	2·550	6·5012	2·554
10	2·796	7·8182	2·732
15	2·998	8·8953	2·955
∞	3·464	12·0000	3·583

Note. For $n=m$ the distribution is symmetrical, with mean at n.

It will be seen that if m and n are large and equal, i.e. $p = q = \frac{1}{2}$ and $Q = pq = \frac{1}{4}$, the limiting values of the second and fourth moments and of the β_2 ratio become 12, 516 and $3\frac{7}{12}$ respectively.

8. Table 4 shows some values of the moment constants for the distribution of d. If these are compared with values given in Tables 2 and 3 it becomes clear that the limiting form has been nowhere nearly reached when $m+n = 16$. Nevertheless there is some similarity in shape, at any rate for m and n not too unequal, which is brought out when the limiting distribution is adjusted to have the same mean and standard deviation as the

distribution of $D_{n,m}$. This is shown in Figs. 1 and 2 in which limiting distributions with $p = \frac{1}{2}$ and $\frac{3}{4}$ have been fitted onto the distributions with $m+n = 16$, $m = 8, 12$. The agreement which can be said still to exist when $m = 12$ has vanished for $m = 14$ (as could be seen were a similar comparison made). At $m = 12$ and 14 the limiting distribution gives an appreciable frequency below zero, when standardized. The adjusted limiting distribution has, of course, smaller intervals between its ordinates than the distribution for finite N, since $\sigma_d > \sigma_D$, and this results in a greater number of ordinates for the former. Thus its frequency polygon is, in general, below that for D. This effect would be corrected if the cumulative distributions were compared.

Table 4. *Moments and standardized deviates for the limiting distribution of d*

Row	$p=m/(m+n)\rightarrow$		0·5	0·6	0·7	0·8	0·9
1	Momental constants for $d=D-m$	Mean	0	−1·67	−3·81	−7·50	−17·78
2		S.D.	3·46	3·70	4·56	6·68	13·57
3		$\sqrt{\beta_1}$	0	−0·49	−0·91	−1·20	− 1·37
4		β_2	3·58	3·92	4·68	5·42	5·87
5	Upper 2·5 % point*	d^+	8 (0·017)	6 (0·015)	4 (0·021)	3 (0·015)	1 (0·012)
6		d^-	7 (0·030)	5 (0·031)	3 (0·050)	2 (0·043)	0 (0·030)
7	Lower 2·5 % point*	d^+	−7 (0·030)	−10 (0·028)	−15 (0·025)	−24 (0·027)	−52 (0·026)
8		d^-	−8 (0·017)	−11 (0·018)	−16 (0·019)	−25 (0·022)	−53 (0·024)
9	Corrected and standardized 2·5 % points	Upper	1·99	1·77	1·57	1·44	1·29
10		Lower	−1·99	−2·20	−2·34	−2·46	− 2·52

* The quantities d^+ and d^- are values of d which may be described as bracketing the 2·5 % point. The figures in parentheses are the probabilities (calculated from Barton's formulae): (i) $P\{d \geqslant d^+\}$ and $P\{d \geqslant d^-\}$ for the upper percentage point, and (ii) $P\{d \leqslant d^+\}$ and $P\{d \leqslant d^-\}$ for the lower percentage point.

9. The following is a possible method of approximating to the percentage points for $D_{n,m}$ when $m+n > 20$, making use of the similarity referred to in the preceding paragraph.

(*a*) First 'standardize' the limiting distribution. The method is illustrated in the case of the 2·5 % points. Rows 5–8 of Table 4 show (for $p = 0{\cdot}5\,(0{\cdot}1)\,0{\cdot}9$) the values of d which bracket the upper and lower 2·5 % points. Linear interpolation between d^+ and d^- gives approximate 2·5 % points to which a continuity correction of 0·5 has been applied. Thus for the lower percentage point when $p = 0{\cdot}8$, $P\{d < -24\} = 0{\cdot}027$ and $P\{d < -25\} = 0{\cdot}022$. Linear interpolation gives $-24{\cdot}40$ as a value for which $P\{d < -24{\cdot}40\} = 0{\cdot}025$ were the distribution continuous. To this value we must add 0·50 as a continuity correction,* and finally have for the standardized deviate

$$\frac{-24{\cdot}40 + 0{\cdot}50 - (-7{\cdot}50)}{6{\cdot}68} = -2{\cdot}46,$$

a figure given in row 10 of the table.

* In most problems of this kind the limiting distribution is continuous and no continuity correction is required to find the standardized limiting percentage point. Here, both the distributions of d and D are discontinuous, but the standardized argument interval of the former does not agree with the unit interval of the latter distribution. The procedure suggested may be described as adding a correction for d and later removing it on a modified scale to get the significance point for D.

(*b*) We now apply to the standardized limits the values of the expectation and standard deviation of D, for any chosen finite values of m and n, giving

$$\text{Percentage limit for } D = \mathscr{E}(D) + \sigma(D) \times \text{standardized limit for } d. \qquad (9)$$

Thus in the case $m+n = 20$, $m = 16$ or $p = 0{\cdot}8$ we have, using Table 4 and the expressions for μ_1' and $\sqrt{\mu_2}$ from equations (4),

$$\text{Upper } 2{\cdot}5\,\% \text{ limit for } D = 10{\cdot}07 + 3{\cdot}994 \times 1{\cdot}44 = 15{\cdot}82.$$

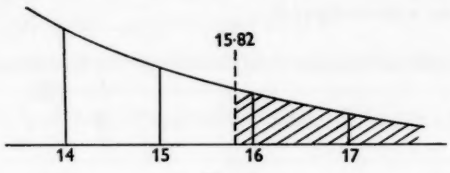


Fig. 3

Table 5. *Comparison of approximate and true 2·5 % limits when* $m+n = 20$

p	m	Standardized limits from Table 4		Case $m+n = 20$		Percentage points using standardizing factors		Limits corrected for discontinuity		True limits from Table 1	
		Upper	Lower	Mean D	σ_D	Upper	Lower	Upper	Lower	Upper	Lower
0·5	10	1·99	−1·99	10·00	2·796	15·56	4·44	17	3	16	4
0·6	12	1·77	−2·20	10·56	2·912	15·71	4·15	17	3	17	4
0·7	14	1·57	−2·34	10·80	3·288	15·96	3·11	17	2	18	3
0·8	16	1·44	−2·46	10·07	3·994	15·82	0·25	17	—	19	1
0·9	18	1·29	−2·52	6·21	4·719	12·30	−5·68	13	—	18	—

Note. The entries '—' in the last column and last column but two indicate that no value of D is significant at the lower 2·5 % level.

If D were continuously distributed, 15·82 would be the approximation to the percentage point; as it is not, and $15{\cdot}82 > 15{\cdot}50$, we cannot include the value $D = 16$ in the $2\frac{1}{2}$ % tail region, but must be content with the statement $P\{D \geqslant 17\} \leqslant 0{\cdot}025$. The position is illustrated in Fig. 3.

10. Proceeding in this way estimates of upper and lower 2·5 % points for the case $m+n = 20$ were obtained as shown in Table 5. The true values, as given in Table 1, are shown for comparison in the final columns. For $p = 0{\cdot}5$, 0·6 and 0·7 the estimates are not more than one unit in error,* but for $p = 0{\cdot}8$ and 0·9 the method is clearly inadequate for $m+n$ as low as 20. Some improvement can be obtained for these extreme values by using the form of limiting distribution suggested by Dr Barton at the end of his note (p. 387), but it is doubtful whether the test should be used at all in samples of this size with $n \leqslant \frac{1}{4}m$.

* In the case $p = 0{\cdot}5$ agreement is really very close; had the entry in col. 7 been 15·50 and that in col. 8 $> 4{\cdot}50$, the figures in cols. 9 and 10 would have been 16 and 4.

11. To sum up, for $m+n>20$ and $m/(m+n)$ not greater than, say, 0·75, approximate limits can be obtained from equation (9), using values of the expectation and standard deviation of $D_{n,m}$ found from the expressions for μ_1' and μ_2 given in equation (4). Values of upper and lower 2·5 % standardized limits for d can be obtained with sufficient accuracy by interpolating in the following table.

Table 6. *Corrected and standardized 2·5 % limits for* $d = D_{n,m}-m$

$p=m/(m+n)$	0·50	0·55	0·60	0·65	0·70	0·75
Upper	1·99	1·88	1·77	1·67	1·57	1·50
Lower	−1·99	−2·10	−2·20	−2·28	−2·34	−2·41

As an example, we may find approximate 2·5 % limits for the case $m = 25$, $n = 15$. Here $p = 0{\cdot}625$ and interpolation in Table 6 gives 1·72, −2·24 as the standardized limits. From equation (4) we find

$$\text{Mean } D_{15,25} = 23{\cdot}03, \quad \mu_2 = 11{\cdot}4236, \quad \text{S.D.} = 3{\cdot}380.$$

This gives as limits for D:

$$23{\cdot}03 + 1{\cdot}72 \times 3{\cdot}380 = 28{\cdot}84 \quad \text{and} \quad 23{\cdot}03 - 2{\cdot}24 \times 3{\cdot}380 = 15{\cdot}46,$$

or, correcting for discontinuity, we should take 30 and 14 as approximations to the upper and lower 2·5 % points. As 15·46 is only just below the borderline value of 15·50 we might reasonably regard 15 rather than 14 as the lower limit.

12. As mentioned in the introduction the Rosenbaum test can be used for the test of the hypothesis which we have set out. His test procedure consists of counting r, the number of observations of one sample which lie outside the extreme points of the other. It is interesting to note that when one sample (say, the m-sample) is wholly included within the extreme values of the other, the Rosenbaum r and D are connected by the relation

$$D_{n,m} = m + r.$$

A comparison of the percentage points of the D-tables and r-tables will show that the upper significance points of $m+r$ are never less than the corresponding significance points of $D_{n,m}$. It might be expected that the D-criterion will be more sensitive to deviations from the hypothesis tested than with the Rosenbaum criterion, but this can only be decided definitely when the power functions of the two tests have been obtained.

Finally I wish to thank Prof. E. S. Pearson and Dr F. N. David for taking a keen interest in these investigations and for a number of helpful suggestions in the preparation of this paper.

REFERENCES

MANN, H. B. & WHITNEY, D. R. (1947). *Ann. Math. Statist.* **18**, 50.
PEARSON, E. S. & HARTLEY, H. D. (1954). Biometrika Tables for Statisticians, **1**. *Cambridge University Press.*
ROSENBAUM, S. (1953). *Ann. Math. Statist.* **24**, 663.

THE LIMITING DISTRIBUTION OF KAMAT'S TEST STATISTIC

Addendum by D. E. BARTON, *University College, London*

The limiting distribution of Kamat's D criterion is easily obtained from the following argument.

Kamat shows that the joint distribution of $i = R_n - n + 1$ and $j = R_m - m + 1$ is given by

$$\left.\begin{aligned} f(i,j) &= 2K && (j = 0 = i) \\ &= K(m-i-1)\,{}^{n+i-2}C_{n-2} && (j = n,\ 0 \leqslant i \leqslant m-1) \\ &= K(n-j-1)\,{}^{m+j-2}C_{m-2} && (0 \leqslant j \leqslant n-1,\ i = m) \\ &= 2K\,{}^{i+j-2}C_{i-1} && (0 \leqslant j \leqslant n-1,\ 0 \leqslant i \leqslant m-1), \end{aligned}\right\}$$

where $$K^{-1} = {}^{n+m}C_m.$$

The marginal distribution of i is consequently

$$f(i) = K(m+1-i)\,{}^{n+i-2}C_{n-2} \quad (0 \leqslant i \leqslant m),$$

for which the mean and variance are

$$\mu_1' = m(n-1)/(n+1), \quad \mu_2 = 2m(n-1)(m+n+1)/\{(n+2)(n+1)^2\}.$$

That of j is the same with n and m interchanged.

If we let $N = m+n$ tend to infinity so that $m/N \to p$, $n/N \to q$, we may write the moments of i as

$$\mu_1' = Np - 2p/q + O(N^{-1}), \quad \mu_2 = 2p/q^2 + O(N^{-1}).$$

Thus if $r = m-i$, $s = n-j$, $d = s-r = D-m$, the first two moments of r and s tend to finite limits and also, since the correlation coefficient of r and s tends to $-\sqrt{(pq)}$, so do the first two moments of d. The probability distribution functions, equally, tend to a proper limit. Thus

$$f(r) = (r+1)\frac{m(m-1)\dots(m-r+1)\,n(n-2)}{N(N-1)\dots(N-r+1)(N-r)(N-r-1)} = (r+1)p^r q^2 + O(N^{-1}),$$

and similarly the probability distribution function of r and s tends to

$$\left.\begin{aligned} f(r,s) &= 2p^{r+1}q^{s+1} && (r>0,\ s>0) \\ &= (s-1)p^2q^s && (r=0,\ s>0) \\ &= (r-1)p^rq^2 && (r>0,\ s=0). \end{aligned}\right\}$$

The probability generating function

$$\pi_{r,s}(x,y) = \Sigma f(r,s)\,x^r y^s = \left\{\frac{pq(x+y-xy)}{(1-px)(1-qy)}\right\}^2$$

follows at once from this, whence, putting

$$x = e^{-it}, \quad y = e^{it},$$

we have the limiting characteristic function of d as

$$\phi_d(t) = \left\{\frac{pq(1-e^{it}+e^{2it})}{(e^{it}-p)(1-qe^{it})}\right\}^2.$$

From this it may be verified that the first moment of d is $2(q-p)/(pq)$ and the higher moments are the same as Kamat's limiting moments of D.

The actual limiting distribution of d is discrete and most simply got from the formula for $f(r,s)$, i.e.

$$\left.\begin{aligned} f(d) &= p^{-d}q^2\{2p^2-(1-Q)(1+d)\}/(1-Q) && (d<0) \\ &= 2Q^2/(1-Q) && (d=0) \\ &= q^dp^2\{2q^2-(1-Q)(1-d)\}/(1-Q) && (d>0), \end{aligned}\right\}$$

where $Q = pq$. This reduces when $p = q = \frac{1}{2}$ to

$$\left.\begin{aligned} f(d) &= \tfrac{1}{6} && (d=0) \\ &= \tfrac{1}{12}(3\,|\,d\,|-1)(\tfrac{1}{2})^{|d|} && (d \neq 0). \end{aligned}\right\}$$

The tail probabilities may be obtained explicitly by summing the expressions for $f(d)$. Thus

$$\text{for } d>0: \quad P(d) = P\{s-r \geqslant d\} = q^d\{pd-1+2q/(1-Q)\},$$

$$\text{for } d<0: \quad P(d) = P\{s-r \leqslant d\} = p^{|d|}\{q\,|\,d\,|-1+2p/(1-Q)\}.$$

These may be used to obtain limiting percentage points.

The modality of the probability distribution function is always triple, as it may be shown that (taking $p<q$ without loss of generality) the following relation holds:

$$\ldots<f(-3)<f(-2)>f(-1)<f(0)>f(1)<f(2)\ldots<f(d_{\max.})>\ldots,$$

where $d_{\max.}$ is the integral part of $1+\{p^{-1}-2q^2/(1-pq)\}$.

The integer $d_{\max.}$ equals 2 if $p>\frac{1}{3}$, but rises rapidly for smaller p (taking the values, 3, 4, 9 as p takes the values 0·3, 0·2, 0·1).

If n is kept finite but $m\to\infty$, a second limiting form may be obtained. In the notation used above

$$\mathscr{E}(s) = O(N^{-1}) = \text{var}\,(s), \quad \mathscr{E}(r) = O(N), \quad \text{var}\,(r) = O(N^2),$$

and hence $x=-d/N$ has the same finite limit as r/m. This is easily seen to be the Pearson Type I curve

$$f(x) = n(n-1)\,x(1-x)^{n-2}.$$

SEQUENTIAL ANALYSIS APPLIED TO CERTAIN EXPERIMENTAL DESIGNS IN THE ANALYSIS OF VARIANCE

BY W. D. RAY

British Coal Utilization Research Association

1. SUMMARY AND INTRODUCTION

On the basis of some fundamental work due to Barnard (1952) and Cox (1952), Johnson (1953) has derived a procedure for applying sequential tests to the general linear hypothesis. Hoel (1955) obtains a similar test by a rather different method.

The general linear hypothesis underlies a number of common analysis of variance situations. It is the purpose of this paper to provide tables which will make it possible to carry out the suggested procedures in the cases of (*a*) one-way classification by groups, and (*b*) randomized blocks. In the course of construction of these tables, we have considered a number of approximations to the confluent hypergeometric function and assessed their accuracy.

Conjectural approximations to the expected sample sizes of the sequential processes considered are also given.

2. A SEQUENTIAL TEST OF THE GENERAL LINEAR HYPOTHESIS

In the classical fixed sample case the general linear hypothesis may be expressed in matrix form as follows:

(i) Let $\mathbf{x} = (x_1, \ldots, x_N)$ be N independent normal variables.

(ii) $\mathscr{E}(\mathbf{x}) = \boldsymbol{\theta}\mathbf{C}'$, where

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s) = (\theta_1, \ldots, \theta_{s-q} \mid \theta_{s-q+1}, \ldots, \theta_s) = (\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}),$$

$$\mathbf{C} = \begin{pmatrix} c_{11} & \cdots & c_{1s} \\ \vdots & & \vdots \\ c_{N1} & \cdots & c_{Ns} \end{pmatrix} = (\mathbf{C}_{(1)}, \mathbf{C}_{(2)}) \quad (s < N),$$

where $\mathbf{C}$ is partitioned similarly to $\boldsymbol{\theta}$.

(iii) $\mathscr{V}(x_i) = \sigma^2$ $(i = 1, \ldots, N)$.

(iv) The θ's and σ are unknown parameters; $\mathbf{C}$ is known and is called the *design matrix*.

The hypothesis to be tested is H_0: $\boldsymbol{\theta}_{(2)} = \mathbf{0}$. The likelihood ratio criterion is then a function of $G = S_b/S_a$, where S_a is the minimum of $(\mathbf{x} - \boldsymbol{\theta}\mathbf{C}')(\mathbf{x} - \boldsymbol{\theta}\mathbf{C}')'$ with respect to $\boldsymbol{\theta}$ and $S_a + S_b$ is the minimum of $(\mathbf{x} - \boldsymbol{\theta}_{(1)}\mathbf{C}'_{(1)})(\mathbf{x} - \boldsymbol{\theta}_{(1)}\mathbf{C}'_{(1)})'$ with respect to $\boldsymbol{\theta}_{(1)}$.

If observations are taken sequentially then at each stage we add a further set of x's, a new row of $\mathbf{C}$ being added for each additional set. Thus after N observations have been taken we may calculate a value $G^{(N)}$, say, of G.

Under certain conditions, which are satisfied in the applications considered in this paper, a sequential test comparing the composite hypotheses

$$H_0: \quad \boldsymbol{\theta}_{(2)} = \mathbf{0},$$

$$H_1: \quad \boldsymbol{\theta}_{(2)}\sigma^{-1} = \boldsymbol{\varphi},$$

may be based on the likelihood ratio (Arnold, 1951)

$$\frac{p(G^{(N)}|\,\boldsymbol{\varphi})}{p(G^{(N)}|\,\mathbf{0})} = e^{-\frac{1}{2}\lambda^{(N)}} M\left(\frac{N-s+q}{2}, \frac{q}{2}; \frac{\frac{1}{2}\lambda^{(N)}G^{(N)}}{1+G^{(N)}}\right), \tag{1}$$

where

$$\left.\begin{aligned} \lambda^{(N)} &= \boldsymbol{\varphi}\mathbf{C}_{(2)}^{(N)\prime}\{\mathbf{I}-\mathbf{C}_{(1)}^{(N)}(\mathbf{C}_{(1)}^{(N)\prime}\mathbf{C}_{(1)}^{(N)})^{-1}\mathbf{C}_{(1)}^{(N)\prime}\}\mathbf{C}_{(2)}^{(N)}\boldsymbol{\varphi}' \\ \text{and} \qquad M(X, Y; u) &= \sum_{j=0}^{\infty} \frac{\Gamma(Y)\,\Gamma(X+j)}{\Gamma(X)\,\Gamma(Y+j)}\frac{u^j}{j!}. \end{aligned}\right\} \tag{2}$$

This procedure is exactly analogous therefore to that employed in comparing two simple hypotheses, i.e.

$$\text{'Accept } H_1 \quad \text{if} \quad \frac{p(G^{(N)}|\,\boldsymbol{\varphi})}{p(G^{(N)}|\,\mathbf{0})} \geqslant \frac{1-\beta}{\alpha}.$$

$$\text{Accept } H_0 \quad \text{if} \quad \frac{p(G^{(N)}|\,\boldsymbol{\varphi})}{p(G^{(N)}|\,\mathbf{0})} \leqslant \frac{\beta}{1-\alpha}.$$

Otherwise take further observations.' α, β are the (approximate) chances of erroneous rejection of H_0, H_1, respectively.

Evidently for all values of $\boldsymbol{\varphi}$ giving the same value for the scalar quantity λ there will be the same sequential procedure.

In the following sections we shall consider the application of this general method to the two special cases mentioned in § 1. In these cases we always have $\lambda = N\delta$, where δ does not depend on N, and so our alternative hypothesis H_1 can be defined in terms of δ.

3. One-way classification by groups

This is the simplest design in the analysis of variance. The data are arranged in k groups, and it is desired to reach conclusions about possible differences between the means of the different groups. The sequential procedure which we shall consider consists of taking an equal number of observations from each group at each stage in the experiment. The theoretical model may be expressed in the following form:

$$x_{ti} = a + b_t + z_{ti} \quad (t = 1, \ldots, k;\ i = 1, \ldots, n),$$

where

$$\mathscr{E}(z_{ti}) = 0, \quad \mathscr{V}(z_{ti}) = \sigma^2, \quad \sum_{t=1}^{k} b_t = 0.$$

In this case $N = kn$, $s = k$ and $q = k-1$:

$$G^{(N)} = \frac{n\sum_{t=1}^{k}(\bar{x}_{t.}-\bar{x}_{..})^2}{\sum_{t=1}^{k}\sum_{i=1}^{n}(x_{ti}-\bar{x}_{t.})^2},$$

$$\lambda^{(N)} = \frac{n\Sigma b_t^2}{\sigma^2} \quad \text{and} \quad \delta = \frac{\Sigma b_t^2}{k\sigma^2}.$$

The procedure is now:

$$\text{'Accept } H_1 \quad \text{if} \quad e^{-\frac{1}{2}\lambda^{(N)}} M\left(\frac{kn-1}{2}, \frac{k-1}{2}; \frac{\frac{1}{2}\lambda^{(N)}G^{(N)}}{1+G^{(N)}}\right) \geqslant \frac{1-\beta}{\alpha},$$

$$\text{Accept } H_0 \quad \text{if} \quad e^{-\frac{1}{2}\lambda^{(N)}} M\left(\frac{kn-1}{2}, \frac{k-1}{2}; \frac{\frac{1}{2}\lambda^{(N)}G^{(N)}}{1+G^{(N)}}\right) \leqslant \frac{\beta}{1-\alpha}.$$

Otherwise take a further set of k (or mk) observations, one (or m) from each group.'

This is equivalent to the conditions: (i) accept H_1 if $G^{(N)} \geqslant \bar{G}^{(N)}$; (ii) accept H_0 if $G^{(N)} \leqslant \underline{G}^{(N)}$; (iii) if $\underline{G}^{(N)} < G^{(N)} < \bar{G}^{(N)}$ continue sampling, where $\underline{G}^{(N)}$, $\bar{G}^{(N)}$ are the solutions of the equations in $G^{(N)}$:

$$\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha} = e^{-\frac{1}{2}\lambda^{(N)}} M\left(\frac{kn-1}{2}, \frac{k-1}{2}; \frac{\frac{1}{2}\lambda^{(N)} G^{(N)}}{1+G^{(N)}}\right).$$

The appropriate limits $\underline{G}^{(N)}$, $\bar{G}^{(N)}$ for $\delta = 0{\cdot}5$, $1{\cdot}0$ and $2{\cdot}0$ with probabilities of error $\alpha = \beta = 0{\cdot}05$ and various values of k are given in Tables 1, 2 and 3 printed on pp. 399–401 below.

4. Randomized blocks

The well-known model for randomized blocks with k 'blocks' and n 'varieties' is

$$x_{ti} = a + b_t + v_i + z_{ti} \quad (t = 1, \ldots, k;\ i = 1, \ldots, r),$$

where

$$\sum_{t=1}^{k} b_t = 0, \quad \sum_{i=1}^{n} v_i = 0$$

and

$$\mathscr{E}(z_{ti}) = 0, \quad \mathscr{V}(z_{ti}) = \sigma^2.$$

The sequential procedure with this design, having fixed the number of varieties n, is to decide at each stage whether or not to add another block of the same n varieties.

Hence we have $N = kn$, $s = k+n-1$ and if H_0 be the hypothesis $v_i = 0$ $(i = 1, \ldots, n)$, then $q = n-1$,

$$G^{(N)} = \frac{k \sum_{i=1}^{n} (\bar{x}_{.i} - \bar{x}_{..})^2}{\sum_{t=1}^{k} \sum_{i=1}^{n} (x_{ti} - \bar{x}_{.i} - \bar{x}_{t.} + \bar{x}_{..})^2},$$

$$\lambda^{(N)} = \frac{k \sum_{i=1}^{n} v_i^2}{\sigma^2} \quad \text{and} \quad \delta = \frac{\sum_{i=1}^{n} v_i^2}{n\sigma^2}.$$

Analogously to the case of the one-way classification by groups, the sequential procedure may now be defined by: (i) accept H_1 if $G^{(N)} \geqslant \bar{G}^{(N)}$, (ii) accept H_0 if $G^{(N)} \leqslant \underline{G}^{(N)}$, (iii) otherwise add another block to the experiment. In this case $\underline{G}^{(N)}$, $\bar{G}^{(N)}$ are the solutions of the following equations in $G^{(N)}$:

$$\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha} = e^{-\frac{1}{2}\lambda^{(N)}} M\left(\frac{k(n-1)}{2}, \frac{n-1}{2}; \frac{\frac{1}{2}\lambda^{(N)} G^{(N)}}{1+G^{(N)}}\right).$$

The appropriate limits $\underline{G}^{(N)}$, $\bar{G}^{(N)}$ for $\delta = 0{\cdot}5$, $1{\cdot}0$ and $2{\cdot}0$ with $\alpha = \beta = 0{\cdot}05$ and various values of n are given in Tables 4, 5 and 6 printed on pp. 402–403 below.

5. Determination of the limits $\underline{G}$, $\bar{G}$

The determination of these limits entails the use of fairly extensive tables of the confluent hypergeometric function $M(X, Y; u)$. In this connexion Rushton (1954) and Rushton & Lang (1954) have provided a fairly comprehensive survey and tabulation of the function. For the values covered by Rushton & Lang's tables, $\underline{G}$ and $\bar{G}$ were calculated by inverse interpolation of the function $\log M(X, Y; u)$ with regard to u.

Unfortunately, the tables were not sufficiently extensive to cover all the cases it was necessary to consider. For example, in the one-way classification by groups for even (odd) k only limits at stages of even (odd) n could be determined, while in the case of randomized

blocks only at odd (even) k for even (odd) n. So, too, it was found for some k and n that not close enough intervals of u had been tabulated to obtain rapid enough convergence of the inverse interpolation process.

In order to cover these situations, attempts were made to obtain useful approximations to the likelihood ratio or equivalently to the confluent hypergeometric function.

6. Approximations to the likelihood ratio

(*a*) The likelihood ratio R is the ratio of a non-central F (or G) distribution to a central F (or G) distribution where the degrees of freedom of the F's are $\nu_1 = q$, $\nu_2 = N-s$ and the non-centrality parameter is λ.

Thus
$$R = \frac{p(F' \mid H_1)_{F'=F}}{p(F \mid H_0)} = \frac{p(G' \mid H_1)_{G'=G}}{p(G \mid H_0)},$$

where $F = \nu_2 G/\nu_1$ and $F' = \nu_2 G'/\nu_1$.

Now
$$p(G' \mid H_1) = \frac{e^{-\frac{1}{2}\lambda}\,\Gamma(\frac{1}{2}(q+N-s))\,G'^{\frac{1}{2}q-1}}{\Gamma(\frac{1}{2}q)\,\Gamma(\frac{1}{2}(N-s))\,(1+G')^{\frac{1}{2}(q+N-s)}}\,M\left(\tfrac{1}{2}(N-s+q), \tfrac{1}{2}q; \frac{\frac{1}{2}\lambda G'}{1+G'}\right),$$

$$p(G \mid H_0) = \frac{\Gamma(\frac{1}{2}(q+N-s))}{\Gamma(\frac{1}{2}q)\,\Gamma(\frac{1}{2}(N-s))}\,\frac{G^{\frac{1}{2}q-1}}{(1+G)^{\frac{1}{2}(q+N-s)}}.$$

Thus
$$R = e^{-\frac{1}{2}\lambda}\,M\left(\tfrac{1}{2}(N-s+q), \tfrac{1}{2}q; \frac{\frac{1}{2}\lambda G}{1+G}\right). \tag{1 \textit{bis}}$$

Patnaik (1949) has approximated to the non-central F-distribution by a central F-distribution having the same first two moments.

In the present instance the F' distribution in the numerator of R would be approximated to by a $kF_{\nu_1\nu_2}$ distribution, where $k = (q+\lambda)/q$, $\nu = (q+\lambda)^2/(q+2\lambda)$. Hence
$$p(G' \mid H_1)_{G'=G} = \frac{\Gamma(\frac{1}{2}(\nu+\nu_2))}{\Gamma(\frac{1}{2}\nu)\,\Gamma(\frac{1}{2}\nu_2)}\left(\frac{\nu}{\nu_1 k}\right)^{\frac{1}{2}\nu}\frac{G'^{\frac{1}{2}\nu-1}}{\left(1+\frac{\nu G'}{\nu_1 k}\right)^{\frac{1}{2}(\nu+\nu_2)}},$$

while
$$p(G \mid H_0) = \frac{\Gamma(\frac{1}{2}(\nu_1+\nu_2))\,G^{\frac{1}{2}\nu_1-1}}{\Gamma(\frac{1}{2}\nu_1)\,\Gamma(\frac{1}{2}\nu_2)\,(1+G)^{\frac{1}{2}(\nu_1+\nu_2)}},$$

and thus
$$R \doteqdot \frac{\Gamma(\frac{1}{2}(\nu+\nu_2))\,\Gamma(\frac{1}{2}\nu_1)}{\Gamma(\frac{1}{2}\nu)\,\Gamma(\frac{1}{2}(\nu_1+\nu_2))}\left(\frac{\nu}{\nu_1 k}\right)^{\frac{1}{2}\nu}\frac{G^{\frac{1}{2}(\nu-\nu_1)}(1+G)^{\frac{1}{2}(\nu_1+\nu_2)}}{\left(1+\frac{\nu G}{\nu_1 k}\right)^{\frac{1}{2}(\nu+\nu_2)}}, \tag{3}$$

where $\nu = (\nu_1+\lambda)^2/(\nu_1+2\lambda)$.

In virtue of (1 *bis*) this can also be regarded as an approximation to the confluent hypergeometric function.

(*b*) While the above approximation (3) holds for all ν_1, ν_2 it was considered worth while to investigate whether any simplification resulted as $\nu_2 \to \infty$. Patnaik (1949) has shown that the first two moments of F' are

$$\mu_1'(F') = \frac{\nu_1+\lambda}{\nu_1}\left(1-\frac{2}{\nu_2}\right)^{-1},$$

$$\mu_2'(F') = \frac{(\nu_1+\lambda)^2+2(\nu_1+2\lambda)}{\nu_1^2}\left(1-\frac{2}{\nu_2}\right)^{-1}\left(1-\frac{4}{\nu_2}\right)^{-1}.$$

Hence for ν_2 large (but not infinite)

$$\mu_2(F') \doteq \frac{2(\nu_1+2\lambda)}{\nu_1^2}\left(1-\frac{2}{\nu_2}\right)^{-2},$$

and therefore $\nu_1(1-2/\nu_2)\,F'$ has moments $\mu_1' = \nu_1+\lambda, \mu_2 = 2(\nu_1+2\lambda)$. We therefore approximate to the distribution of F' by regarding $\nu_1(1-2/\nu_2)\,F'$ as being distributed approximately as χ'^2 with ν_1 degrees of freedom and non-centrality parameter λ.

Inserting the corresponding approximations for the distribution of G, we obtain

$$\begin{aligned} R &\doteq e^{-\frac{1}{2}\lambda}\sum_{r=0}^{\infty}\frac{\Gamma(\frac{1}{2}\nu_1)}{\Gamma(\frac{1}{2}\nu_1+r)}\frac{\{\frac{1}{4}\lambda(\nu_2-2)\,G\}^r}{r!} \\ &= e^{-\frac{1}{2}\lambda}\frac{\Gamma(\frac{1}{2}\nu_1)\,I_{\frac{1}{2}\nu_1-1}(\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}})}{(\frac{1}{2}\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}})^{\frac{1}{2}\nu_1-1}}, \end{aligned} \tag{4}$$

where $I_{\frac{1}{2}\nu_1-1}$ is a Bessel function of imaginary argument.

When $\nu_1 = 1$, that is, for example, when $k = 2$ in the one-way classification by groups, we have

$$\begin{aligned} R &\doteq \frac{e^{-\frac{1}{2}\lambda}\,\Gamma(\frac{1}{2})\,I_{-\frac{1}{2}}(\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}})}{(\frac{1}{2}\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}})^{-\frac{1}{2}}} \\ &= e^{-\frac{1}{2}\lambda}\cosh\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}}. \end{aligned} \tag{5}$$

This may be compared with an approximation to the confluent hypergeometric function given by Arnold (1951).

Using the relationship

$$I_X(\zeta) = \frac{(\frac{1}{2}\zeta)^X}{\Gamma(X+1)}e^{\zeta}\,M(X+\tfrac{1}{2}, 2X+1; -2\zeta),$$

and Kummer's formula $M(X, Y; z) = e^z\,M(Y-X, Y; -z)$, we obtain (4) in the form

$$R \doteq e^{-\frac{1}{2}\lambda}\,e^{-\{\lambda(\nu_2-2)G\}^{\frac{1}{2}}}\,M(\tfrac{1}{2}\nu_1-\tfrac{1}{2}, \nu_1-1; 2\{\lambda(\nu_2-2)\,G\}^{\frac{1}{2}}).$$

Comparing (1 *bis*) and (4) it may be noted that we have in fact approximated to a confluent hypergeometric function by a similar function of a slightly different form.

7. Construction of the tables of limits $\underline{G}$, $\overline{G}$

In order to decide on the extent of tabulation required, the results of Baker (1950) were taken as a rough guide. In an empirical investigation into the distribution of sample size in sequential tests comparing certain simple hypotheses he found that in less than 1 in 20 cases did the actual observed sample size exceed three times the expected sample size. The tables were accordingly constructed to cover sample sizes up to three times the expected sample size, and it is to be anticipated that they will cover most of the cases encountered in practice.

Wherever possible the logarithm of the expression (1 *bis*) was equated in turn to $\log(1-\beta)/\alpha$, $\log\beta/(1-\alpha)$, and the solutions $\overline{G}$, $\underline{G}$ respectively obtained by inverse interpolation. By far the greater content of Tables 1–6 was arrived at by this method. To produce certain intermediate values when, in the case of the one-way classification, n was small or when the range of the tables of the confluent hypergeometric function was exceeded, recourse was made to approximation (3). Solutions $\underline{G}'$, $\overline{G}'$ obtained by the Newton–Raphson iterative method applied to approximation (3) were of great assistance in approximating

to the exact values. It was found that $\bar{G}'$ was a very close estimator of $\bar{G}$ but that $\underline{G}'$ was not equivalently as good an estimator of $\underline{G}$, at least for small n (or k in the case of randomized blocks). However, it was found that the rate of change of the limits for increasing n (or k) was almost identical using the correct form (1 *bis*) or the approximate one (3). Thus knowledge of the approximate limits and their differences for increasing n (or k) could be utilized to obtain good estimates of the true limits providing a few exact values had already been obtained for small n (or k). Estimates obtained this way are printed in italic type in Tables 1–6.

Table 7. *Comparison of limits obtained from equations* (1 *bis*) *and* (3)

One-way classification by groups. $\delta=1$, $\alpha=\beta=0{\cdot}05$.

$k=2$

n	λ	$\underline{G}$	$\bar{G}$	$\underline{G}'$	$\bar{G}'$
4	8	0·052	2·390	0·096	2·410
6	12	0·105	1·016	0·138	1·001
8	16	0·135	0·710	0·164	0·700
10	20	0·150	0·578	0·179	0·580
12	24	0·166	0·504	0·195	0·502

$k=5$

n	λ	$\underline{G}$	$\bar{G}$	$\underline{G}'$	$\bar{G}'$
3	15	0·331	2·469	0·348	2·497
5	25	0·340	0·973	0·357	0·958
7	35	0·331	0·687	0·347	0·685
9	45	0·322	0·565	0·338	0·568
11	55	0·314	0·498	0·330	0·503

$\underline{G}$, $\bar{G}$ were obtained from (1 *bis*); $\underline{G}'$, $\bar{G}'$ from (3).

Table 8. *Comparison of limits obtained from equations* (1 *bis*) *and* (5)

One-way classification by groups. $\delta=0{\cdot}5$, $\alpha=\beta=0{\cdot}05$.

$k=2$

n	λ	$\underline{G}$	$\bar{G}$	$\underline{G}''$	$\bar{G}''$
6	6	0·002	1·091	0·012	0·918
8	8	0·025	0·639	0·032	0·608
10	10	0·040	0·471	0·047	0·466
12	12	0·050	0·385	0·058	0·387
14	14	0·059	0·333	0·067	0·337
16	16	0·066	0·300	0·074	0·302
18	18	0·072	0·276	0·079	0·277
20	20	0·076	0·258	0·083	0·258
30	30	0·089	0·205	0·097	0·207

$\underline{G}$, $\bar{G}$ were obtained from (1 *bis*); $\underline{G}''$, $\bar{G}''$ from (5).

The approximation (4) has not been investigated numerically, except for the case when $\nu_1 = 1$. It is found that it gives reasonable agreement of the limits when n is fairly large. Once again, however, the rate of change of limits with increasing n (or k) is similar whether (1 *bis*) or (5) is used to obtain them.

Thus either approximation may be used together with some known exact limits obtained from (1 *bis*) to expand or interpolate where required in Tables 1–6. (Tables 7 and 8 give a few examples of the accuracy of these approximations.)

It may be noted from Tables 1–6 that in certain cases no decision is possible before a certain stage. This is because G by definition must be real, finite and positive. Thus in some practical situations an initial sample must be taken before the sequential procedure can be put into operation.

8. The practical use of the tables

Example 1. Suppose we require to take observations from each of three groups and postulate that the model of § 3 might describe the situation. We wish to test the null hypothesis that each group mean is zero against the alternative hypothesis that

$$\delta = \frac{\sum_{i=1}^{3} b_i^2}{3\sigma^2} = 0{\cdot}5.$$

The probabilities of error α, β are taken to be equal to 0·05.

Table 9

n	x_{1i}	Σx_{1i}	Σx_{1i}^2	x_{2i}	Σx_{2i}	Σx_{2i}^2	x_{3i}	Σx_{3i}	Σx_{3i}^2	Sum of squares		G
										Between	Within	
1	0·6	0·6	0·36	2·3	2·3	5·29	−7·8	−7·8	60·84	—	—	—
2	−4·4	−3·8	19·72	13·4	15·7	184·85	9·3	1·5	147·33	—	—	—
3	15·1	11·3	247·73	3·8	19·5	199·29	13·4	14·9	326·89	—	—	—
4	2·8	14·1	255·57	5·7	25·2	231·78	−6·1	8·8	364·10	—	—	—
5	−6·1	8·0	292·78	9·9	35·1	329·79	8·1	16·9	429·71	76·32	735·96	0·104
6	3·0	11·0	301·78	5·7	40·8	362·28	10·1	27·0	531·72	74·14	776·67	0·095
7	0·6	11·6	302·14	−1·1	39·7	363·49	2·9	29·9	540·13	58·12	833·67	0·070

Three samples from normal populations whose means are zero and whose standard deviations are 10 have been chosen to illustrate the procedure.

In Table 9 the approximate calculations are shown, a value of G being calculated at each stage of sampling after $n = 5$. Table 1 is consulted until one or other of the corresponding limits is violated. The example considered results in a correct decision in favour of the null hypothesis when $n = 7$.

Example 2. A practical application in which the randomized block design of § 4 may be used, occurs when sampling for dust concentration in a rectangular duct (see Fig. 1). Sampling probes can be transversely inserted in four positions P_1, P_2, P_3, P_4 and collections of dust made over the width of the duct. It is required to determine whether or not the effect of gravity on the dust flow is creating stratification from top to bottom. Sampling is carried

out at each of the four positions in random order repeating the scheme after a series of fixed time intervals. The accuracy of measurement of dust concentration is known beforehand.

In this situation the 'varieties' correspond to the four positions and the 'blocks' to the elements of the time sequence. The hypothesis of stratification is numerically interpreted as a particular value of δ and the sampling procedure carried out until a decision is reached on its validity.

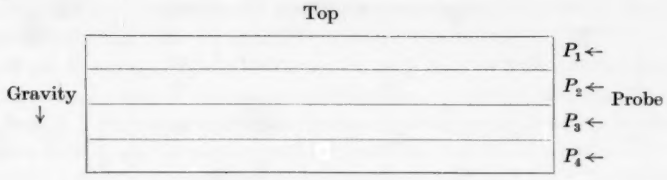


Fig. 1. Cross-section of rectangular duct.

9. EXPECTED SAMPLE NUMBER

Although no general formula for the expected sample number for composite hypotheses exists, Bhate (1954) conjectures that the natural generalization of Wald's formula:

$$\mathscr{E}[\log R \mid N = \mathscr{E}(N)] \doteqdot \begin{cases} (1-\alpha)\log\dfrac{\beta}{1-\alpha} + \alpha\log\dfrac{1-\beta}{\alpha} & \text{if } H_0 \text{ is true,} \\ \beta\log\dfrac{\beta}{1-\alpha} + (1-\beta)\log\dfrac{1-\beta}{\alpha} & \text{if } H_1 \text{ is true,} \end{cases} \tag{6}$$

where R is the likelihood ratio, may give useful results.

As a first approximation, using the method described below, the left-hand side of the equation was evaluated as the value of $\log R$ when $N = \mathscr{E}(N)$. The values so obtained are, of course, both conjectural and approximate. In Table 10 (*a*) and (*b*) they are compared with the sample size required for the (fixed sample) analysis of variance test having the same prescribed chances of error as the sequential test. The latter were found from the graphs of the power functions of analysis of variance tests given by Pearson & Hartley (1951).

It will be seen that, as in other cases of sequential testing of hypotheses, the apparent saving in sample size is of the order of one-third to one-half of the fixed sample size.

Substituting the appropriate expression for R from (1 *bis*), the left-hand side of equation (6) is

$$\mathscr{E}[-\tfrac{1}{2}\lambda\log_{10}e + \log_{10}M(X, Y; u)],$$

where $$X = \tfrac{1}{2}(N-s+q), \quad Y = \tfrac{1}{2}q, \quad u = \tfrac{1}{2}\lambda G/(1+G).$$

Now by Taylor's theorem

$$\log_{10}M(X, Y; u) = \log_{10}M(X, Y; \mathscr{E}(u)) + \frac{(u-\mathscr{E}(u))}{1!}\frac{\partial}{\partial u}(\log_{10}M)\Big|_{u=\mathscr{E}(u)} + \frac{(u-\mathscr{E}(u))^2}{2!}\frac{\partial^2}{\partial u^2}(\log_{10}M)\Big|_{u=\mathscr{E}(u)} + \ldots.$$

Thence we have to a first approximation

$$-\tfrac{1}{2}\lambda\log_{10}e + \log_{10}M(X, Y; \mathscr{E}(u)).$$

When H_0 is true it may be shown that

$$\mathscr{E}\left(\frac{G}{1+G}\right) = \frac{Y}{X} \quad \text{and when } H_1 \text{ is true} \quad \mathscr{E}\left(\frac{G}{1+G}\right) = e^{-\frac{1}{2}\lambda} \sum_{m=0}^{\infty} \frac{(\frac{1}{2}\lambda)^m (Y+m)}{m!\,(X+m)}.$$

Table 10. *Conjectural expected sample sizes (n or k) for the sequential test when H_0 is true and corresponding values for the fixed sample size test, when $\alpha = \beta = 0{\cdot}05$*

(*a*) One-way classification by groups

	$\delta = 0{\cdot}5$		$\delta = 1$		$\delta = 2$	
	n		n		n	
k	Sequential	Fixed	Sequential	Fixed	Sequential	Fixed
2	10	14	5	8	3	4–5
3	8	11	4	6	2	4
4	6–7	9	3	5	2	4
5	6	8	3	4	2	3
6	4	7	3	4	2	3
7	4	6	3	4	.	.
8	.	.	2–3	4	.	.
9	.	.	2	3	.	.

(*b*) Randomized blocks

	$\delta = 0{\cdot}5$		$\delta = 1$		$\delta = 2$	
	k		k		k	
n	Sequential	Fixed	Sequential	Fixed	Sequential	Fixed
2	11	15	5–6	9	3	6
3	8	12	4	7	2–3	4–5
4	7	10	4	6	2	4
5	6	8–9	3	5	2	3
6	5	8	3	4–5	2	3
7	5	7	3	4	.	.
8	4	6	2–3	4	.	.

Thus on the assumption that H_0 is true we have, for $\alpha = \beta = 0{\cdot}05$:

(i) For the one-way classification by groups, to solve for n the equation

$$-\tfrac{1}{2}kn\delta \log_{10} e + \log_{10} M\{\tfrac{1}{2}(kn-1), \tfrac{1}{2}(k-1); \tfrac{1}{2}(k-1)\,\delta\} \doteqdot -1{\cdot}15,$$

where in reaching this result $nk/(nk-1)$ has been taken as unity.

(ii) For the case of randomized blocks, to solve for k the equation

$$-\tfrac{1}{2}kn\delta \log_{10} e + \log_{10} M\{\tfrac{1}{2}(k(n-1)),\ \tfrac{1}{2}(n-1);\ \tfrac{1}{2}n\delta\} \doteqdot -1{\cdot}15.$$

Such solutions are those tabulated in Table 10 (a) and (b). Examination of further terms in the Taylor expansion reveal that no wide discrepancy results from only considering its first term.

Several sampling experiments have been carried out and give average sample sizes in very close agreement with those obtained by the above method.

This conjectural formula of Bhate's can also be used to give the expected sample size in a two-sided sequential t-test. This test is equivalent to a one-way classification by groups when $k = 2$. Arnold (1951) reports the sample sizes at which a decision was reached for each of 500 such tests both when the null hypothesis was true and when a particular alternative was true. The δ of Arnold's paper is, in our terminology, $\sqrt{(\frac{1}{2}\delta)}$ and his v^2 corresponds to our

$$nG/(1+G) = \frac{\frac{1}{2}n^2(\bar{x}_{1.} - \bar{x}_{2.})^2}{\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2}.$$

The probabilities of error were taken to be $\alpha = \beta = 0{\cdot}05$ and $\delta = 1$.

The average sample number when the null hypothesis was true was reported as 10·03, while the value given by the conjectural formula was $N = 2n = 10{\cdot}0$.

On the assumption that H_1 is true we have to determine

$$\mathscr{E}\left(\frac{G}{1+G}\right) = e^{-\frac{1}{2}\lambda} \sum_{m=0}^{\infty} \frac{(\frac{1}{2}\lambda)^m (Y+m)}{m!\,(X+m)}$$

for different values of X, Y.

A more amenable form of this expression may be obtained thus:

Let $$S = e^{-\frac{1}{2}\lambda} \sum_{m=0}^{\infty} \frac{(\frac{1}{2}\lambda)^m (Y+m)}{m!\,(X+m)} = e^{-\frac{1}{2}\lambda}(Y-X)T + 1,$$

where $$T = \frac{1}{X} + \frac{\frac{1}{2}\lambda}{1!\,(X+1)} + \frac{(\frac{1}{2}\lambda)^2}{2!\,(X+2)} + \ldots.$$

Now $$(\tfrac{1}{2}\lambda)^X T = \int_0^{\frac{1}{2}\lambda} (\tfrac{1}{2}\lambda)^{X-1} e^{\frac{1}{2}\lambda}\, d(\tfrac{1}{2}\lambda) = \psi_X, \quad \text{say,}$$

and a reduction formula for ψ_X is

$$\psi_X = (\tfrac{1}{2}\lambda)^{X-1} e^{\frac{1}{2}\lambda} - (X-1)\psi_{X-1}.$$

Alternatively $$\psi_X = e^{\frac{1}{2}\lambda}[(\tfrac{1}{2}\lambda)^r - r(\tfrac{1}{2}\lambda)^{r-1} + r(r-1)(\tfrac{1}{2}\lambda)^{r-2} \ldots (-1)^r r!].$$

Hence $$S = \frac{e^{-\frac{1}{2}\lambda}(Y-X)}{(\frac{1}{2}\lambda)^X}\psi_X + 1.$$

A few calculations to determine the expected sample size when the alternative hypothesis H_1 is true gave results almost identical to those obtained when H_0 is true.

10. Comments

As far as is known there is no proof that tests of this type terminate with probability one. Recently, however, David and Kruskal (1956) have provided such a proof for the sequential t-test. His proof uses some asymptotic forms for large n of the numerator and denominator of the likelihood ratio. It may be therefore that the asymptotic forms (4, 5) of § 6 may be helpful in proving an analogous result for the tests described in this paper.

It is perhaps worth while noting that other designs in the analysis of variance may find the following tables applicable. For example, in a factorial experiment it may be required to know whether or not first to take two levels of a factor, then three or four until some hypothesis is accepted or rejected. So too in a hierarchal classification by groups a series of successive subgroups within main groups could be added until some postulated hypothesis was accepted or denied.

The question of choosing the sequential test appropriate to fit a given situation is never altogether easy. In the present case if k (or n) is fixed by the nature of the problem, we have still a choice of values for the parameters α, β and δ, and in making this choice we may well be influenced by the expected sample size.* The tables given in the present paper are, of course, far from complete as far as α, β and δ are concerned, but it is hoped that they will give enough information to make some application of theory to practice possible.

I would like to thank Dr N. L. Johnson most gratefully for his advice and encouragement in the preparation of this paper and the Director General of the British Coal Utilization Research Association for permission to publish it.

REFERENCES

Arnold, K. J. (1951). Introduction to *Tables to facilitate sequential t-tests*. National Bureau of Standards, A.S.M. **7**.

Baker, A. G. (1950). Properties of some tests in sequential analysis. *Biometrika*, **37**, 334.

Barnard, G. A. (1952). The frequency justification of certain sequential tests. *Biometrika*, **39**, 144.

Bhate, D. H. (1954). Ph.D. Thesis, University of London.

Cox, D. R. (1952). Sequential tests for composite hypotheses. *Proc. Camb. Phil. Soc.* **48**, 290.

David, H. T. & Kruskal, W. H. (1956). The WAGR sequential t-test reaches a decision with probability one. *Ann. Math. Statist.* **27**, 797.

Hoel, P. G. (1955). On a sequential test for the general linear hypothesis. *Ann. Math. Statist.* **26**, 136.

Johnson, N. L. (1953). Some notes on the application of sequential methods in the analysis of variance. *Ann. Math. Statist.* **24**, 614.

Patnaik, P. B. (1949). The non-central χ^2 and F distributions and their applications. *Biometrika*, **36**, 202.

Pearson, E. S. & Hartley, H. O. (1951). Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. *Biometrika*, **38**, 112.

Rushton, S. (1954). On the confluent hypergeometric function $M(\alpha, \gamma; x)$. *Sankhyā*, **13**, 369.

Rushton, S. & Lang, E. D. (1954). Tables of the confluent hypergeometric distribution. *Sankhyā*, **13**, 377.

* Considerations that arise in the interpretation of δ were discussed in another connexion by Pearson & Hartley (1951, pp. 125–9) when considering the use of charts for the power function of analysis of variance tests.

Table 1. *One-way classification by groups (true limits)*

k=no. of groups; n=no. within a group; $\lambda = nk\delta$; $\delta = 0{\cdot}5$; $\alpha = \beta = 0{\cdot}05$.

$k=2$				$k=3$				$k=4$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
4	4	—	5·390	5	7·5	0·037	1·319	4	8	0·065	1·825
6	6	0·002	1·091	7	10·5	·072	0·696	6	12	·110	0·770
8	8	·025	0·639	9	13·5	·089	·497	8	16	·126	·521
10	10	·040	·471	11	16·5	·099	·400	10	20	·131	·411
12	12	·050	·385	13	19·5	·105	·344	12	24	·134	·350
14	14	0·059	0·333	15	22·5	0·108	0·306	14	28	0·136	0·310
16	16	·066	·300	17	25·5	·110	·280	16	32	·138	·282
18	18	·072	·276	19	28·5	·112	·261				
20	20	·076	·258	21	31·5	·115	·245				
30	30	0·089	0·205								

$k=5$				$k=6$				$k=7$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
3	7·5	0·072	4·176	2	6	0·008	—	3	10·5	0·184	2·407
5	12·5	·142	0·927	4	12	·168	1·272	5	17·5	·211	0·784
7	17·5	·155	·568	6	18	·183	0·646	7	24·5	·207	·512
9	22·5	·158	·432	8	24	·183	·464	9	31·5	·199	·401
11	27·5	·159	·360	10	30	·180	·377				
13	32·5	0·159	0·317	12	36	0·176	0·327				
15	37·5	·157	·287								

Table 2. *One-way classification by groups.*
(*Approximate adjusted values are printed in italics*)

k = no. of groups; n = no. within a group; $\lambda = nk\delta$; $\delta = 1{\cdot}0$; $\alpha = \beta = 0{\cdot}05$.

$k=2$				$k=3$				$k=4$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
4	8	0·052	2·390	3	9	0·124	4·760	2	8	0·144	—
5	10	*·082*	*1·380*	4	12	*·169*	*1·79*	3	12	*·231*	*3·13*
6	12	·105	1·016	5	15	·195	1·174	4	16	·266	1·498
7	14	*·121*	*0·826*	6	18	*·210*	*0·902*	5	20	*·276*	*1·040*
8	16	·135	·710	7	21	·221	·762	6	24	·284	0·838
10	20	0·150	0·578	9	27	0·234	0·605	8	32	0·287	0·637
12	24	·166	·504	11	33	·240	·522	10	40	·286	·540
16	32	·187	·424	13	39	·244	·471	12	48	·284	·482
20	40	·199	·383	15	45	·247	·436	16	64	·280	*·424*
								20	80	·277	*·391*
30	60	0·215	0·333	21	63	0·251	0·377				
60	120	*·235*	*·300*	31	93	*·251*	*·334*	30	120	*0·271*	*0·342*
				51	153	*·251*	*·306*	50	200	*·266*	*·306*

$k=5$				$k=6$				$k=7$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
3	15	0·331	2·469	2	12	0·381	24·042	3	21	0·469	1·925
4	20	*·345*	*1·332*	3	18	*·405*	*2·133*	4	28	*·447*	*1·175*
5	25	·340	0·973	4	24	·398	1·237	5	35	·420	0·888
6	30	*·335*	*·792*	5	30	*·384*	*0·920*	6	42	*·405*	*·745*
7	35	·331	·687	6	36	·373	·763	7	49	·387	*·645*
9	45	0·322	0·565	8	48	0·354	0·601	9	63	0·365	*0·552*
11	55	·314	·498	10	60	·340	*·521*	11	77	·349	*·494*
13	65	·308	*·456*	12	72	·330	*·473*	13	91	*·338*	*·456*
15	75	·303	*·428*	14	84	*·322*	*·440*	15	105	*·328*	*·430*
25	125	*0·288*	*0·360*	20	120	*0·307*	*0·386*	21	147	*0·315*	*0·384*

$k=8$				$k=9$				$k=10$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
2	16	0·566	7·500	3	27	0·565	1·684	2	20	0·709	5·066
3	24	*·518*	*1·792*	4	36	*·510*	*1·085*	4	40	·534	*1·06*
4	32	·479	1·120	5	45	·470	*0·845*	6	60	·457	*0·705*
5	40	*·448*	*0·865*					8	80	*·415*	*·575*
6	48	·424	*·730*	7	63	0·421	*0·635*	10	100	*·389*	*·506*
				9	81	*·392*	*·540*				
8	64	0·391	*0·585*	11	99	*·372*	*·485*	16	160	*0·347*	*0·414*
10	80	*·371*	*·512*					20	200	*·332*	*·384*
16	128	*·336*	*·414*	15	135	*0·347*	*0·425*				
20	160	*·324*	*·387*	21	189	*·325*	*·380*	30	300	*0·312*	*0·348*
				31	279	*·306*	*·345*	40	400	*·300*	*·329*
30	240	*0·306*	*0·348*	41	369	*·296*	*·328*				
40	400	*·288*	*·318*								

Table 3. *One-way classification by groups.*

(*Approximate adjusted values are printed in italics*)

k = no. of groups; n = no. within a group; $\lambda = nk\delta$; $\delta = 2{\cdot}0$; $\alpha = \beta = 0{\cdot}05$.

$k=2$				$k=3$				$k=4$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
2	8	0·121	—	3	18	0·458	3·361	2	16	0·612	22·175
4	16	·279	2·153	5	30	·497	1·403	4	32	·616	*1·696*
6	24	·346	1·232	7	42	·507	*1·040*	6	48	·596	*1·128*
8	32	·380	0·967	9	54	·510	*0·891*	8	64	*·583*	*0·935*
10	40	·401	·842	11	66	*·512*	·809	10	80	*·574*	·837
12	48	0·416	*0·775*	13	78	*0·512*	*0·758*	12	96	*0·567*	*0·777*
16	64	*·434*	*·700*	15	90	*·511*	*·724*	16	128	*·557*	*·708*
20	80	*·444*	*·660*					20	160	*·550*	*·669*
				21	126	*0·510*	*0·661*				
30	120	*0·461*	*0·611*	31	186	*·507*	*·611*	30	240	*0·540*	*0·619*
40	160	*·479*	*·580*	41	246	*·505*	*·592*	40	320	*·535*	*·596*
50	200	*·474*	*·572*	51	306	*·504*	*·578*	50	400	*·532*	*·582*

$k=5$				$k=6$			
n	λ	$\underline{G}$	$\bar{G}$	n	λ	$\underline{G}$	$\bar{G}$
3	30	0·747	*2·445*	2	24	0·950	—
5	50	*·677*	*1·272*	4	48	*·765*	*1·535*
7	70	*·637*	*0·994*	6	72	*·685*	*1·081*
9	90	*·613*	*·869*	8	96	*·645*	*0·913*
11	110	*·596*	*·799*	10	120	*·620*	*·825*
13	130	*0·584*	*0·753*	12	144	*0·602*	*0·770*
15	150	*·575*	*·720*	16	192	*·578*	*·706*
				20	240	*·563*	*·669*
21	210	*0·556*	*0·662*				
31	310	*·540*	*·617*	30	360	*0·542*	*0·621*
41	410	*·531*	*·595*	40	480	*·530*	*·598*
51	510	*·525*	*·582*	50	600	*·523*	*·583*

Table 4. *Randomized blocks (true limits)*

k = no. of blocks; n = no. of varieties; $\lambda = nk\delta$; $\delta = 0{\cdot}5$; $\alpha = \beta = 0{\cdot}05$.

$n=2$				$n=3$				$n=4$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
5	5	—	186·97	4	6	0·005	11·366	3	6	0·006	—
7	7	0·027	2·586	6	9	·082	1·637	5	10	·120	1·635
9	9	·061	1·403	8	12	·117	0·953	7	14	·155	0·876
11	11	·086	1·009	10	15	·137	·706	9	18	·168	·634
								11	22	·175	·516
13	13	0·105	0·814	12	18	0·148	0·580				
15	15	·120	·698	14	21	·155	·503	13	26	0·180	0·446
17	17	·133	·620	16	24	·160	·451	15	30	·182	·400
								17	34	·183	·367
27	27	0·170	0·448	22	33	0·174	0·365	19	38	·184	·343

$n=5$				$n=6$				$n=7$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
4	10	0·145	2·129	3	9	0·151	4·680	2	7	0·089	—
6	15	·181	0·914	5	15	·211	1·061	4	14	·236	1·419
8	20	·194	·627	7	21	·218	0·658	6	21	·286	0·731
10	25	·198	·499	9	27	·216	·505	8	28	·234	·529
				11	33	·212	·424	10	35	·226	·433
12	30	0·197	0·426								
14	35	·196	·380	13	39	0·208	0·374				
16	40	·194	·347								

$n=8$			
k	λ	$\underline{G}$	$\bar{G}$
3	12	0·254	2·627
5	20	·268	0·870
7	28	·254	·573
9	36	·241	·452

Table 5. *Randomized blocks.* (*Approximate adjusted values are printed in italics*)

k=no. of blocks; n=no. of varieties; $\lambda = nk\delta$; $\delta = 1{\cdot}0$; $\alpha = \beta = 0{\cdot}05$.

$n=2$				$n=3$				$n=4$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
3	6	0·006	—	2	6	0·009	—	3	12	0·293	*6·32*
5	10	·147	5·383	4	12	·236	3·774	5	20	·359	*1·5*
7	14	·218	2·129	6	18	·302	1·529	7	28	·373	*0·987*
9	18	·269	1·470	8	24	·331	1·071	9	36	·375	*·795*
11	22	·304	1·188	10	30	·346	0·874	11	44	·375	*·694*
13	26	0·329	1·032	12	36	0·355	0·765				
15	30	·349	0·933								

$n=5$				$n=6$				$n=7$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
2	10	0·308	—	3	18	0·470	*2·86*	2	14	0·529	2·748
4	20	·412	*1·803*	5	30	·453	*1·137*	4	28	·507	1·412
6	30	·411	*1·02*	7	42	·430	*0·812*	6	42	·463	0·878
8	40	·402	*0·780*	9	54	·412	*·675*				

$n=8$			
k	λ	$\underline{G}$	$\bar{G}$
3	24	0·581	2·15
5	40	·506	*0·999*
7	56	·460	*·740*

Table 6. *Randomized blocks.* (*Approximate adjusted values are printed in italics*)

k=no. of blocks; n=no. of varieties; $\lambda = nk\delta$; $\delta = 2{\cdot}0$; $\alpha = \beta = 0{\cdot}05$.

$n=2$				$n=3$				$n=4$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
3	12	0·365	—	2	12	0·488	—	3	24	0·779	4·482
5	20	·557	*4·379*	4	24	·672	3·399	5	40	·779	*1·86*
7	28	·655	*2·519*	6	36	·716	*1·9*	7	56	*·766*	*1·39*
9	36	·715	*1·977*	8	48	*·734*	*1·49*	9	72	*·754*	*1·19*
11	44	·756	*1·720*	10	60	*·742*	*1·299*	11	88	*·745*	*1·08*
15	60	*0·806*	*1·47*								

$n=5$				$n=6$			
k	λ	$\underline{G}$	$\bar{G}$	k	λ	$\underline{G}$	$\bar{G}$
2	20	0·920	29·864	3	36	0·978	*2·904*
4	40	·858	*2·102*	5	60	*·884*	*1·517*
6	60	*·810*	*1·406*	7	84	*·830*	*1·189*
8	80	*·780*	*1·166*	9	108	*·800*	*1·040*

LOGNORMAL APPROXIMATION TO PRODUCTS AND QUOTIENTS

By S. R. BROADBENT

The British Coal Utilization Research Association

SUMMARY. Measurements have been made which are subject to error, and we are required to give limits to a combination of the values measured. The combinations considered here are products and quotients. Some exact results are available in simple cases, but otherwise an approximation is required. The lognormal distribution, which is asymptotically exact, is shown to give useful approximations when fitted by moments to the combination. This method of fitting is nearly optimum in a defined metric. Tables are given which make its application simple.

1. Introduction

The problem we shall consider is assessing the precision of certain combinations of measurements which have been made with a known distribution of error. Suppose the unbiased measurements x_i $(i=1, \ldots, n)$ are made, that $E[x_i] = \mu_i$, and that we are interested in a combination of the x_i of the form

$$q = (x_1 x_2 \ldots x_j)/(x_{j+1} \ldots x_n) \quad (1 \leqslant j \leqslant n)$$

(more generally, the observations x_i may be raised to any power).

We may be required to set fiducial limits to

$$(\mu_1 \mu_2 \ldots \mu_j)/(\mu_{j+1} \ldots \mu_n),$$

i.e. to combine the fiducial distributions of the μ_i which we have attempted to measure, or to set probability limits to q, i.e. to combine the probability distributions of the x_i. In the first case we assume the distributions of x_i known, except for the means μ_i, and that the observations x_i are given; in the second we assume the distributions of the x_i completely known. The two problems are formally identical,* and we shall henceforward speak only of setting limits to q.

To fix the ideas with a particular example of the first case, consider the efficiency E of a steam boiler determined by a single trial in which the heat supplied to and the heat obtained from the boiler are obtained from sample measurements. The efficiency may be calculated from

$$E = HW(1-C)/\{Qh(1-C')\},$$

where H is the heat in 1 lb. of steam, W the weight of water evaporated during the trial, C the air-dried moisture content of the coal, Q the weight of coal fired, h its air-dried calorific value and C' its moisture content as fired. All these measurements are subject to errors whose distribution is known; the effect of these errors on the precision of E is required.

In general, the error of a measurement has two possible sources: sampling technique and instrument. In this paper we shall for simplicity suppose the errors to be either normal or rectangular, of known coefficients of variation or half-range, and with small coefficients of variation, although the restriction to these distributions is not necessary. In many applications, the distributions of the errors are known to be of one of these two forms, and their standard deviations or half-ranges can usually be found by simple investigations. In other applications the errors can only be guessed, and then the roughest approximations are appropriate.

* We assume that fiducial distributions can be combined in the same way as probability distributions, but see the discussion in Creasy (1954) on this point.

It is impossible to list exhaustively all such combinations of errors. Some of the simpler combinations can be discussed individually; in the event of a large number of errors being combined we can use with confidence the asymptotic distribution. It is in the intermediate cases that approximations must be critically considered. The choice of a suitable family of approximating distributions will always be more of an art than a science.

It is possible in a particular case to refine approximation to any required degree. Gram–Charlier or similar series may be used; the saddlepoint method given by Daniels (1954) also generates approximating functions. However, in this paper we are concerned with a working or first-order approximation only: all that some data and applications merit.

It is well known that, even when n is small and the component distributions are far from normal, $x_1+\ldots+x_n$ has a distribution close to normal. When the component variates are not added but multiplied and divided, the fundamental approximating distribution is the lognormal, as was pointed out, for example, by Shellard (1952). We consider the questions, how is this distribution best fitted and when may we use it with confidence?

2. KNOWN RESULTS

2·1. We now suppose the errors of measurement are independent, and we denote a normal variate by N and a rectangular variate by R. The quotient of the standard deviation and the mean of N, and the quotient of the half-range and mean of R are both denoted by α_1, α_2, etc., referring to the variates $x_1, x_2, \ldots$ in this order. Thus R/N ($\alpha_1 = 0{\cdot}1$, $\alpha_2 = 0{\cdot}05$) denotes the quotient of a rectangular variate whose mean is ten times its half-range and an independent normal variate whose coefficient of variation is 5 %.

For $q = N/N$ Geary's approximation (1930) is used in all practical cases. This states that if Q is the quotient of the means of the numerator and denominator,

$$(q-Q)/(\alpha_1^2 q^2 + \alpha_2^2 Q^2)^{\frac{1}{2}}$$

is approximately normally distributed with zero mean and unit variance. This approximation is very good up to values of α_1 and α_2 as large as 0·25 (see Creasy, 1954). The exact percentage points of this distribution may also be calculated using existing tables as was shown by Fieller (1932). The product $N \times N$ was discussed by Craig (1936) and Aroian (1947).

Percentage points of R/N have been given by the author (1954). He has given also points of the quotient of a triangular and a normal variate.

The distributions of $R \times R$, R/R and so on are not difficult to calculate exactly.

Distributions of products, quotients and powers of variates have been extensively studied, but not many results have been obtained that can be applied to practical problems.

3. LOGNORMAL APPROXIMATIONS

3·1. The distribution of

$$q = (x_1 x_2 \ldots x_j)/(x_{j+1} \ldots x_n)$$

tends to the lognormal as $n \to \infty$ under very general conditions. The lognormal distribution has been discussed by Finney (1941), Gaddum (1945) and Johnson (1949).

The most general lognormal approximation to q is a variate z such that $\log(z-\xi)$ is normally distributed with mean μ and variance σ^2. Here for simplicity we consider the choice of μ and σ only, ξ being always taken as zero. A variate with this distribution is necessarily positive, while q may have negative values. When $\alpha_1, \alpha_2, \ldots$ are small, the approximation by this lognormal distribution may nevertheless be satisfactory, since the probability attached to such negative values is very small.

3·2. The first method of choosing μ and σ is to calculate the moments of $\log q$ and to set μ and σ^2 equal to the first and second corrected moments. We call this the *method of fitting by moments to* $\log q$, or the *fit to* $\log q$; this method is used for estimation by Finney (1941). It is also intuitively attractive to choose that lognormal approximation whose mean and variance are equal to the mean and variance of q (see Wicksell, 1917). We call this the *method of fitting by moments to* q, or the *fit to* q. The two fits are in general different, although the difference tends to zero as n increases. The difference throws a doubt on the intuitive appeal of fitting by moments, and raises the question we consider below, by what criterion are we to choose our approximation?

Let a lognormal distribution have mean m and variance s^2, and let the distribution of the logarithm of a variate from the lognormal have mean μ and variance σ^2. Then it is known that

$$\left.\begin{aligned} m &= \exp(\mu+\sigma^2/2), \\ s^2 &= \{\exp(2\mu+\sigma^2)\}\{\exp(\sigma^2)-1\}, \\ \mu &= \log m - \tfrac{1}{2}\log(1+s^2/m^2), \\ \sigma^2 &= \log(1+s^2/m^2). \end{aligned}\right\} \tag{1}$$

For the fit to $\log q$, we must find the mean and variance of $\log q$, and set μ and σ^2 equal to them. The $t\,\%$ point of this fit is

$$\exp(\mu+p_t\sigma), \tag{2}$$

where the probability is $t/100$ that a standardized normal variate is less than p_t.

For the fit to q, we must find the mean and variance of q, and set m and s^2 equal to them. The $t\,\%$ point of this fit is

$$(m\exp[p_t\{\log(1+y^2)\}^{\frac{1}{2}}])/(1+v^2)^{\frac{1}{2}}, \tag{3}$$

where $100v = 100s/m$, the coefficient of variation of q. This $t\,\%$ point is tabulated in Table 1 for $m = 1$, $100v = 0(0{\cdot}01)15$ and $t = 1$, 5, 95 and 99. To use this table it is necessary to find the mean and coefficient of variation of q, to enter the table with the appropriate v, and multiply the value in the table by m. The table was calculated from a series expansion of (3).

3·3. To find the moments of q or of $\log q$ we require the moments of various powers (positive and negative) of x or of $\log x$, when x is normally and when x is rectangularly distributed. We write $x = \mu(1+\alpha y)$, where y is either a standardized normal variate or is uniformly distributed between -1 and 1, and α is small (less than 0·15). We begin by finding the moments of $\log(1+\alpha y)$, not when y is normally distributed, but (since $\Pr\{1+\alpha y<0\} = \epsilon$ is less than 10^{-10}) when y is from the truncated distribution

$$\exp(-\tfrac{1}{2}y^2)\,dy/\{(1-\epsilon)\sqrt{(2\pi)}\} \quad (y > -1/\alpha), \tag{4}$$

which is for practical purposes indistinguishable from the normal. We avoid mathematical difficulties by taking this distribution as the parent. Alternatively, the treatment of convergence given by Derksen (1939) could be used.

The moment-generating function of $\log(1+\alpha y)$ is $E[(1+\alpha y)^{it}]$, and this may be written

$$\frac{1}{\sqrt{(2\pi)}\,(1-\epsilon)}\int_{-1/\alpha}^{1,\alpha}(1+it\alpha y+\tfrac{1}{2}it(it-1)\alpha^2y^2+\ldots)\exp(-\tfrac{1}{2}y^2)\,dy+K,$$

where $|K| \leqslant \epsilon/(1-\epsilon)$.

Let

$$\begin{aligned} J(\alpha,r) &= \frac{1}{\sqrt{(2\pi)}}\int_{1/\alpha}^{\infty} y^r\exp(-\tfrac{1}{2}y^2)\,dy \\ &= \{1-I[(1/(2\alpha^2)),\tfrac{1}{2}(r-1)]\}\,2^{\frac{1}{2}r-1}\Gamma[\tfrac{1}{2}(r+1)]/\sqrt{\pi}. \end{aligned}$$

Here $I(y, p)$ is the Incomplete Gamma ratio $\int_0^y v^p e^{-v}\, dv/\Gamma(p+1)$, and is nearly one when y is large and p small; we deduce from Pearson's table (1922) that for $\alpha \leqslant 0{\cdot}15$ and $r \leqslant 6$, $J(\alpha, r) \leqslant 10^{-6}$, and we therefore neglect it in this region.

The series above is uniformly convergent within the range of integration, so that term-by-term integration is permissible. We may replace the limits of integration by $(-\infty, \infty)$ in the first four non-zero terms when $\alpha \leqslant 0{\cdot}15$, since the error so introduced is a factor of $\{1 - 2J(\alpha, r)/(1-\epsilon)\}$ for $r = 0$, 2, 4 and 6. As K is negligible we obtain to sufficient accuracy the first four terms of the convergent series for the moment-generating function

$$1 + \alpha^2 it(it-1)/2 + \alpha^4 it(it-1)(it-2)(it-3)/8 + \alpha^6 it \ldots (it-5)/48,$$

and hence the cumulants of $\log(1+\alpha y)$,

$$\left.\begin{aligned} \mu_1' = \kappa_1 &= -\alpha^2/2 - 3\alpha^4/4 - 5\alpha^6/2 - \ldots, \\ \mu_2 = \kappa_2 &= \alpha^2 + 5\alpha^4/2 + 32\alpha^6/3 + \ldots, \\ \kappa_3 &= -3\alpha^4 - 22\alpha^6 - \ldots, \\ \kappa_4 &= 20\alpha^6 + \ldots. \end{aligned}\right\} \tag{5}$$

The cumulants of $\log(1+\alpha y)$ when y is rectangularly distributed in $(-1, 1)$ are obtained in a similar way. We have

$$\left.\begin{aligned} \mu_1' = \kappa_1 &= -\alpha^2/6 - \alpha^4/20 - \alpha^6/42 - \ldots, \\ \mu_2 = \kappa_2 &= \alpha^2/3 + 7\alpha^4/45 + 29\alpha^6/315 + \ldots, \\ \kappa_3 &= -2\alpha^4/15 - 1513\alpha^6/3780 - \ldots, \\ \kappa_4 &= -2\alpha^4/15 - 16\alpha^6/315 - \ldots. \end{aligned}\right\} \tag{6}$$

We next require the expectation of various powers of $(1+\alpha y)$ when y is normally distributed and when y is rectangularly distributed. These have been calculated for the truncated normal distribution (4) and the rectangular distribution; the arguments are similar to those given above and are not repeated here. Finally, we have arranged the results commonly required in Table 2. These formulae, with (2) or Table 1, enable the percentage points of the two lognormal fits to q to be calculated.

The power series above have been formed by identifying coefficients in the expansion of the moment-generating function. We have given only the first four terms of this expansion, for moderate α. For later terms or larger α, important corrections would have to be applied to the coefficients obtained by uncritically continuing the series. Indeed, Wicksell (1921) pointed out that some of the series in (5) and (6) may diverge if the later coefficients are uncritically formed. The simple rules for obtaining these coefficients apply only under the conditions stated.

The results given may be extended to cases in which the variates are not independent. It is necessary only to find the moments for correlated variates in the same way as above; some moments are given by Haldane (1942).

As an example, consider the lognormal approximations to $q = N/N$; by convention the coefficients of variation of numerator and denominator are respectively $100\alpha_1$ and $100\alpha_2$. Using Table 2 we obtain

$$E[q] = m = (1)(1 + \alpha_2^2 + 3\alpha_2^4 + \ldots),$$

and

$$E[q^2] = (1+\alpha_1^2)(1 + 3\alpha_2^2 + 15\alpha_2^4 + \ldots),$$

i.e.

$$V[q] = s^2 = \alpha_1^2 + \alpha_2^2 + 3\alpha_1^2\alpha_2^2 + 8\alpha_2^4 + \ldots.$$

Table 1. *Standardized lognormal percentage points*

Given the mean m and standard deviation s, let $v = 100s/m$. The percentage point of the lognormal distribution with this mean and standard deviation is the entry in the table, multiplied by m.

Lower 1 % points ($t = 1$)

v	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	Δ
0	1·0000	9977	9954	9930	9907	9884	9861	9838	9815	9792	−23
1	0·9770	9747	9724	9701	9679	9656	9633	9611	9588	9566	−23
2	0·9544	9521	9499	9477	9454	9432	9410	9388	9366	9344	−22
3	0·9322	9300	9278	9256	9234	9213	9191	9169	9148	9126	−22
4	0·9105	9083	9061	9040	9019	8997	8976	8955	8933	8912	−21
5	0·8892	8870	8849	8828	8807	8787	8766	8745	8724	8703	−21
6	0·8683	8662	8641	8621	8600	8580	8559	8539	8519	8498	−20
7	0·8478	8458	8438	8418	8398	8378	8357	8338	8318	8298	−20
8	0·8278	8258	8238	8219	8199	8179	8160	8140	8121	8101	−20
9	0·8082	8062	8043	8024	8004	7985	7966	7947	7928	7909	−19
10	0·7890	7871	7852	7833	7814	7795	7776	7758	7739	7720	−19
11	0·7702	7683	7665	7646	7628	7609	7591	7572	7554	7536	−18
12	0·7518	7500	7481	7462	7445	7427	7409	7391	7374	7356	−18
13	0·7338	7320	7302	7285	7267	7249	7232	7214	7197	7179	−17
14	0·7162	7144	7127	7110	7093	7075	7058	7041	7024	7007	−17
15	0·6990	—	—	—	—	—	—	—	—	—	—

Lower 5 % points ($t = 5$)

v	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	Δ
0	1·0000	9983	9967	9951	9934	9918	9902	9885	9869	9853	−16
1	0·9836	9820	9804	9788	9771	9755	9739	9723	9707	9691	−16
2	0·9675	9658	9642	9626	9610	9594	9578	9562	9546	9530	−16
3	0·9514	9498	9482	9467	9451	9435	9419	9403	9388	9372	−16
4	0·9356	9340	9324	9309	9293	9278	9262	9246	9231	9215	−16
5	0·9200	9184	9168	9153	9137	9122	9106	9091	9075	9060	−15
6	0·9045	9029	9014	8999	8983	8968	8953	8937	8922	8907	−15
7	0·8892	8877	8862	8846	8831	8816	8801	8786	8771	8756	−15
8	0·8741	8726	8711	8696	8681	8666	8651	8636	8622	8607	−15
9	0·8592	8577	8562	8548	8533	8518	8503	8489	8474	8460	−15
10	0·8445	8430	8415	8401	8386	8372	8357	8343	8328	8314	−15
11	0·8299	8285	8271	8256	8242	8228	8213	8199	8185	8170	−14
12	0·8156	8142	8128	8113	8099	8085	8071	8057	8043	8029	−14
13	0·8015	8001	7987	7973	7959	7945	7931	7917	7903	7889	−14
14	0·7875	7861	7848	7834	7820	7806	7793	7779	7765	7751	−14
15	0·7738	—	—	—	—	—	—	—	—	—	—

Table 1 (*cont.*)

Upper 5% points ($t = 95$)

v	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	Δ
0	1·0000	0016	0033	0049	0066	0083	0099	0115	0132	0149	+17
1	1·0165	0182	0198	0215	0232	0249	0265	0282	0299	0316	+17
2	1·0332	0349	0366	0383	0400	0417	0433	0450	0467	0484	+17
3	1·0501	0518	0535	0552	0569	0586	0603	0620	0637	0654	+17
4	1·0671	0688	0705	0723	0740	0757	0774	0791	0809	0826	+17
5	1·0843	0860	0878	0895	0912	0930	0947	0964	0982	0999	+17
6	1·1017	1034	1051	1069	1086	1104	1121	1139	1156	1174	+17
7	1·1191	1209	1226	1244	1262	1279	1297	1315	1332	1350	+18
8	1·1368	1385	1403	1421	1439	1456	1474	1492	1510	1528	+18
9	1·1545	1563	1581	1599	1617	1635	1653	1671	1689	1707	+18
10	1·1725	1742	1760	1779	1797	1815	1833	1851	1869	1887	+18
11	1·1905	1923	1941	1959	1978	1996	2014	2032	2050	2069	+18
12	1·2087	2105	2123	2142	2160	2178	2196	2215	2233	2252	+18
13	1·2270	2288	2307	2325	2343	2362	2380	2399	2417	2436	+18
14	1·2454	2472	2491	2510	2528	2547	2565	2584	2602	2621	+19
15	1·2639	—	—	—	—	—	—	—	—	—	—

Upper 1% points ($t = 99$)

v	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	Δ
0	1·0000	0023	0047	0070	0094	0117	0141	0164	0188	0211	+23
1	1·0235	0259	0282	0306	0330	0354	0378	0402	0426	0450	+24
2	1·0474	0498	0522	0547	0571	0595	0620	0644	0669	0693	+24
3	1·0718	0743	0767	0792	0817	0841	0866	0891	0916	0941	+25
4	1·0966	0991	1016	1041	1067	1092	1117	1142	1168	1193	+25
5	1·1219	1244	1279	1296	1321	1347	1372	1398	1424	1450	+26
6	1·1476	1502	1528	1554	1580	1606	1632	1659	1685	1711	+26
7	1·1735	1764	1790	1817	1843	1870	1896	1923	1950	1977	+27
8	1·2004	2030	2057	2084	2111	2138	2165	2193	2220	2247	+27
9	1·2274	2302	2329	2356	2384	2411	2439	2466	2494	2522	+28
10	1·2549	2577	2605	2633	2661	2689	2716	2745	2773	2801	+28
11	1·2829	2857	2885	2914	2942	2970	2999	3027	3056	3084	+28
12	1·3113	3141	3170	3199	3228	3257	3285	3314	3343	3372	+29
13	1·3402	3431	3460	3489	3518	3547	3577	3606	3636	3665	+29
14	1·3695	3724	3753	3785	3813	3843	3872	3902	3932	3962	+30
15	1·3992	—	—	—	—	—	—	—	—	—	—

Similarly, $E[\log q] = \mu = -(\frac{1}{2}\alpha_1^2 + \frac{3}{4}\alpha_1^4 + \ldots) + (\frac{1}{2}\alpha_2^2 + \frac{3}{4}\alpha_2^4 + \ldots)$,

and $V[\log q] = \sigma^2 = (\alpha_1^2 + \frac{5}{2}\alpha_1^4 + \ldots) + (\alpha_2^2 + \frac{5}{2}\alpha_2^4 + \ldots)$.

Using these values of μ and σ^2 we obtain for the first two moments of the lognormal fitted to $\log q$,

mean: $1 + \alpha_2^2 + \frac{1}{2}\alpha_1^4 + \frac{5}{2}\alpha_2^4 + \ldots$,

variance: $\alpha_1^2 + \alpha_2^2 + 3\alpha_1^4 + 3\alpha_1^2\alpha_2^2 + 5\alpha_2^4 + \ldots$.

Since these agree with m and s^2 to $O(\alpha_1^2, \alpha_2^2)$, and to higher order when $\alpha_1 = \alpha_2$, the difference between the two fits will be small.

Table 2

	r	y is distributed normally with mean zero and variance one	y is distributed rectangularly in the interval $(-1, 1)$
$E[(1+\alpha y)^r]$	$\frac{1}{2}$	$1-\alpha^2/8-15\alpha^4/128-\ldots$	$1-\alpha^2/24-\alpha^4/128-\ldots$
	1	1	1
	2	$1+\alpha^2$	$1+\alpha^2/3$
	4	$1+6\alpha^2+3\alpha^4$	$1+2\alpha^2+\alpha^4/5$
	$-\frac{1}{2}$	$1+3\alpha^2/8+105\alpha^4/128+\ldots$	$1+\alpha^2/8+7\alpha^4/128+\ldots$
	-1	$1+\alpha^2/+3\alpha^4+\ldots$	$1+\alpha^2/3+\alpha^4/5+\ldots$
	-2	$1+3\alpha^2+15\alpha^4+\ldots$	$1/(1-\alpha^2)$
	-4	$1+10\alpha^2+105\alpha^4+\ldots$	$1+10\alpha^2/3+7\alpha^4+\ldots$
$E[\log(1+\alpha y)^r]$		$-r(\alpha^2/2+3\alpha^4/4+\ldots)$	$-r(\alpha^2/6+\alpha^4/20+\ldots)$
$V[\log(1+\alpha y)^r]$		$r^2(\alpha^2+5\alpha^4/2+\ldots)$	$r^2(\alpha^2/3+7\alpha^4/45+\ldots)$

It is not possible to state how different the two will be in general. We compare below the percentage points given by the two fits with some exact values (Table 3). The comparisons are made for quite large coefficients of variation, and for distributions rather different from lognormal. The agreement is surprisingly good (it would be worse at more extreme percentage points). The two fits generally give points closer to each other than to the exact value, i.e. choice between the two methods does not appear to be important.

3·4. We turn now to the *normal approximation*. Cramér (1951) has shown that if H is a function of the central moments of a multi-dimensional sample of size S, such that H and its first two derivatives with respect to these moments are continuous, then H is asymptotically normal. If $H(m_\nu, \ldots)$ is this function, its mean and variance are asymptotically $H(\mu_\nu, \ldots)$, and

$$\mu_2(m_\nu)\left(\frac{\partial H}{\partial \mu_\nu}\right)^2 + \ldots + 2\mu_{11}(m_\nu m_\lambda)\left(\frac{\partial H}{\partial \mu_\nu}\right)\left(\frac{\partial H}{\partial \mu_\lambda}\right) + \ldots.$$

Now $q = (x_1 x_2 \ldots x_j)/(x_{j+1} \ldots x_n)$ is in the form required by this theorem, with $S = 1$. It is sometimes said that q is approximately normally distributed, with mean

$$(\mu_1\mu_2 \ldots \mu_j)/(\mu_{j+1} \ldots \mu_n),$$

and variance $$\sigma_1^2\left(\frac{\partial q}{\partial x_1}\right)^2 + \ldots + 2\rho_{12}\sigma_1\sigma_2\left(\frac{\partial q}{\partial x_1}\right)\left(\frac{\partial q}{\partial x_2}\right) + \ldots,$$

that is, with coefficient of variation

$$100\{\alpha_1^2 + \ldots + \alpha_{j+1}^2 + \ldots + 2\rho_{12}\alpha_1\alpha_2 + \ldots - 2\rho_{1,j+1}\alpha_1\alpha_{j+1} - \ldots\}^{\frac{1}{2}}.$$

The mean and variance of q agree with these formulae to $O(\alpha^2)$ and $O(\alpha^4)$ respectively when the x_i are normally distributed, and so do not differ greatly for small α. Now the normal and lognormal distributions do not differ greatly in their 1, 5, 95 and 99 % points for small coefficients of variation. If the lognormal approximation to q is good, the normal approximation will also be good for small α. This appears to be a better justification for the normal approximation in such cases than reliance on the application of Cramér's asymptotic theorem to a sample of size one. Brunt (1931) has derived this approximation by a Taylor expansion of q and the assumption of normality.

4. General use of the approximation

4·1. It is necessary to know how useful the lognormal approximation is in practice. Since the exact distributions for which we require approximate percentage points are not usually known, it is impossible to give exact results. Equally, in many problems it is as reasonable to suppose the variates are lognormally distributed as to consider them normally distributed, and then the question is trivial.

In Table 3 some typical comparisons with simple distributions are given. It is intuitively clear that the approximation will improve as the α decrease or as the number of component variates increases. If we are satisfied by the agreement indicated in Table 3, we may use the method with confidence in more complicated situations with smaller α.

But as the number of variates increases, larger coefficients of variation than those given in Table 3 may be allowed without impairing the approximation. It is interesting to know how much larger the coefficients of variation may be. Some quantitative conclusions may be drawn from the cumulants of $\log q$, where q is the product of n independent normal variates:

$$q = x_1 x_2 \dots x_n.$$

Suppose the constant $\mu_1 \mu_2 \dots \mu_n$ has already been removed, so that x_i has mean one and variance α_i^2: suppose also that the lognormal approximation to q has been deemed satisfactory. We now wish to approximate to $q' = qx_{n+1}$, where x_{n+1} has mean one and variance β^2. The lognormal approximation to q' will remain satisfactory for small β; for large β the distribution of q' will be unduly affected by x_{n+1} and the lognormal approximation will no longer satisfy us. The problem is to determine conditions on β, in relation to the α_i, which allow us to use with confidence a new lognormal approximation. These conditions cannot ensure that the new approximation is from every point of view as good as the old, for example, the exact and approximate cumulative distributions will not generally coincide at the points where they coincided before.

It follows from (5) that the fit to $\log q$ supposes that

$$u = (\log q - \kappa_1)/\kappa_2^{\frac{1}{2}}$$

is approximately normally distributed with zero mean and unit variance, where

$$\kappa_1 = -\lambda/2 - 3\mu/4 - 5\nu/2,$$

and

$$\kappa_2 = \lambda + 5\mu/2 + 32\nu/3.$$

Here

$$\lambda = \sum_{i=1}^{n} \alpha_i^2, \quad \mu = \sum_{i=1}^{n} \alpha_i^4 \quad \text{and} \quad \nu = \sum_{i=1}^{n} \alpha_i^6.$$

The third and fourth cumulants of u are approximately

$$\kappa_3 = -3\mu/\lambda^{\frac{3}{2}} \quad \text{and} \quad \kappa_4 = 20\nu/\lambda^2,$$

and are of the order of $3\alpha/n^{\frac{1}{2}}$ and $20\alpha^2/n$ respectively.

For q' similar relations hold, with

$$\lambda' = \lambda + \beta^2, \quad \mu' = \mu + \beta^4 \quad \text{and} \quad \nu' = \nu + \beta^6.$$

If the new third cumulant is less than or equal to κ_3 in absolute value, and the new fourth cumulant less than or equal to κ_4, we have grounds for believing that the normal approximation to $\log q'$ and the lognormal approximation to q' will be at least as satisfactory as the approximations to q. The condition on the third cumulant is approximately

$$\lambda^3(\mu+\beta^4)^2 \leqslant \mu^2(\lambda+\beta^2)^3.$$

Table 3. *Exact and lognormal approximation percentage points*

Distribution ...		*R*	*N*	*R*/*R*	*R*/*N*	*N* × *N*	*N*/*N*
$100\alpha_1$ ...		20	10	10	10	4	5
$100\alpha_2$ ...		—	—	10	5	4	5
1 %	Exact	**0·804**	**0·767**	**0·842**	**0·846**	**0·872***	**0·848**
	Fit to q	0·760	0·789	0·827	0·837	0·875	0·848
	Fit to $\log q$	0·758	0·786	0·827	0·836	0·877	0·848
5 %	Exact	**0·820**	**0·836**	**0·872**	**0·882**	**0·908***	**0·890**
	Fit to q	0·822	0·845	0·874	0·881	0·910	0·890
	Fit to $\log q$	0·820	0·842	0·874	0·881	0·910	0·890
95 %	Exact	**1·180**	**1·164**	**1·147**	**1·133**	**1·095***	**1·124**
	Fit to q	1·200	1·173	1·144	1·134	1·096	1·124
	Fit to $\log q$	1·203	1·175	1·144	1·134	1·096	1·124
99 %	Exact	**1·196**	**1·233**	**1·188**	**1·185**	**1·135***	**1·180**
	Fit to q	1·298	1·255	1·209	1·194	1·139	1·179
	Fit to $\log q$	1·303	1·259	1·210	1·195	1·140	1·180

* Exact points communicated by Prof. L. A. Aroian.

We write $b = \beta^2/\lambda$ and $\gamma = \lambda^2/\mu$; then the condition on b is

$$b^3\gamma^2 - b^2 + (2\gamma - 3)\,b - 3 \leqslant 0.$$

It follows from Cauchy's inequality that $1 \leqslant \gamma \leqslant n$. The condition is satisfied at $b = 0$ and until

$$\gamma = \{(b+1)^{\frac{3}{2}} - 1\}/b^2.$$

Similarly, the condition on the fourth cumulant is approximately

$$\lambda^2(\nu+\beta^6) \leqslant (\lambda+\beta^2)^2\,\nu.$$

We write $\delta = \lambda^3/\nu$; it follows from Cauchy's inequality that $1 \leqslant \delta \leqslant n^2$. The condition becomes

$$\delta b^2 - b - 2 \leqslant 0,$$

and is satisfied at $b = 0$ and until

$$b = \{(1+8\delta)^{\frac{1}{2}} + 1\}/2\delta.$$

Similar regions may be defined when the x_i are raised to positive or negative powers, and for rectangular variates. When q consists of the product of independent rectangular variates we see from (6) that both κ_3 and κ_4 are $O(\alpha^4)$, and hence that the conditions on both these cumulants result in an acceptable region similar to the first of the two defined above.

4·2. Some practical conclusions are now drawn.

All coefficients of variation equal. Suppose $\alpha_1 = \ldots = \alpha_n = \alpha$ and x_{n+1} has coefficient of variation β. Then the lognormal fit to the new product will be as good as the fit to the old if β is less than the value given in Table 4, $n = 1\,(1)\,6$; the asymptotic criterion is also shown.

Table 4. *Critical values of β for $\alpha_1 = \ldots = \alpha_n = \alpha$*

n	Criterion κ_3	κ_4
1	1·46α	1·41α
2	1·38α	1·30α
3	1·32α	1·26α
4	1·29α	1·24α
5	1·26α	1·24α
6	1·23α	1·23α
∞	1·22α	1·19α

Table 5. *Critical values of β, α_i unequal*

n	α_2	α_3	Criterion κ_3	κ_4
2	$\alpha/2$	—	1·28α	1·29α
2	$\alpha/5$	—	1·40α	1·39α
3	$\alpha/2$	$\alpha/2$	1·27α	1·25α
3	$\alpha/5$	$\alpha/5$	1·21α	1·36α

Coefficients of variation unequal. For various values of α_2 and α_3, Table 5 gives the critical values of β in terms of $\alpha_1 = \alpha$.

It is clear from Tables 4 and 5 that when a new component is added to q and it is known that the lognormal approximation to q is satisfactory, the coefficient of variation of the new variate can be of the order of 1·25 times the largest coefficient of variation already present. When this is the case, we have some confidence that the new q' will also have a satisfactory lognormal approximation. In this way we may extend the results of Table 3. We may, for example, conclude that it is likely that the lognormal approximations are satisfactory for RN/N, $\alpha_1 \leqslant 0{\cdot}1$, $\alpha_2 \leqslant 0{\cdot}06$ and $\alpha_3 \leqslant 0{\cdot}05$, and for RN/NN, $\alpha_1 \leqslant 0{\cdot}1$, $\alpha_2 \leqslant 0{\cdot}06$, $\alpha_3 \leqslant 0{\cdot}07$ and $\alpha_4 \leqslant 0{\cdot}05$.

5. THE DISTANCE BETWEEN DISTRIBUTIONS

5·1. We have already pointed out that there are several methods of choosing a lognormal approximation to q, and we have given examples to show that the fit by moments to q gives good practical results. It is possible to give this method some theoretical justification, which of course rests on the criterion adopted for judging an approximation.

Suppose that we are given the cumulative distribution function $F(x)$, for which the first and second moments μ and σ^2 exist. Suppose also the cumulative distribution $G(x)$ has moments m and s^2, and that we are to choose from some family that $G(x)$ which most resembles $F(x)$. For example, $G(x)$ may be a well-tabulated distribution, and we are at liberty to choose the parameters m and s^2. The manner in which $G(x)$ is to resemble $F(x)$ is of course critical, and the number of possible criteria is unbounded. We develop first two almost trivial criteria and then discuss two approaches to the distance between $F(x)$ and $G(x)$.

5·2. In the first place, suppose we really require two percentage points of $F(x)$; this is the type of problem we stated initially. Suppose two parameters of $G(x)$ are available, and for particular values of these parameters the two percentage points of $G(x)$ coincide with those of $F(x)$. These are the values we should choose. For example, in fitting a normal distribution to the 5 and 95 % of R/N, $\alpha_1 = 0{\cdot}1$, $\alpha_2 = 0{\cdot}05$, we have to solve

$$m - 1{\cdot}645s = 0{\cdot}882, \quad m + 1{\cdot}645s = 1{\cdot}133,$$

to obtain a unique solution. This method cannot be usefully applied, since if the points were known we would not require to fit $G(x)$.

Now suppose it is known that a good fit will be obtained by selecting a set of known points for agreement. For example, the author (1954) gives the 1 and 5 % points of R/N; suppose the $2\frac{1}{2}$ % point is required. The lognormal distribution is a reasonable fit to this distribution; we therefore fit a normal distribution to the logarithms of the 1 and 5 % points p_1 and p_5. We solve

$$m - 2{\cdot}326s = \log p_1, \quad m - 1{\cdot}645s = \log p_5.$$

The $2\frac{1}{2}$ % point, $p_{2{\cdot}5}$, is then given by

$$m - 1{\cdot}960s = 0{\cdot}462 \log p_1 + 0{\cdot}538 \log p_5 = \log p_{2{\cdot}5}.$$

Calculations show that this interpolation is within the accuracy of the tables given. The method is also exact when applied to interpolation or extrapolation in Table 1 of this paper.

5·3. The distance between two cumulative distribution functions at the value x is naturally defined as some monotonic increasing function, $M(z)$, of $|F(x) - G(x)|$. The distance between the two functions may then be defined by either

$$\int_{-\infty}^{\infty} M(|F(x) - G(x)|)\, \psi\{F(x)\}\, dF(x),$$

or

$$\sup_x [M(|F(x) - G(x)|)\, \psi\{F(x)\}],$$

where $\psi(z)$ is some weighting function. The first definition is associated with Cramér (1928), von Mises (1931) and Smirnov (1936); the second with Kolmogorov (1933).

We are concerned with distances along the x-axis rather than perpendicular to it. That is, we aim to fit approximate percentage points which are close to the true points rather than to fix points at which $F(x)$ is close to the desired value.

Suppose that $F(x) = p$ and $G(x) = p$ have respectively the unique inverses $x = \phi(p)$ and $x = \gamma(p)$ for almost all p. We are concerned with a distance between the functions at p of $M(|\phi(p) - \gamma(p)|)$, and we require definitions of distance between the distributions in the form

$$\int_0^1 M(|\phi(p) - \gamma(p)|)\, \psi(p)\, dp,$$

or

$$\sup_p [M(|\phi(p) - \gamma(p)|)\, \psi(p)].$$

We do not here consider the second definition, which gives an unbounded distance for certain very simple cases, e.g. if $F(x)$ and $G(x)$ are normal distributions differing only in variance, $M(z) = z$ and $\psi(p) = 1$.

A natural form of the first definition for our purpose is given by $M(z) = z^2$, and $\psi(p)$ the characteristic function of some set E contained in the interval $(0, 1)$ in which we are specially interested. For example, E might be the interval $(0, 1)$ itself, or the union of the intervals $(0{\cdot}001, 0{\cdot}05)$ and $(0{\cdot}95, 0{\cdot}995)$.

We therefore define the distance θ between $F(x)$ and $G(x)$ by

$$\theta = \int_E \{\phi(p) - \gamma(p)\}^2 \, dp / |E|.$$

The standardized form of x corresponding to $F(x)$ is $(x-\mu)/\sigma$; let $\xi(p)$ be the inverse of the distribution function corresponding to the standardized form and $\eta(p)$ the similar function derived from $G(x)$:

$$\phi(p) = \mu + \sigma\xi(p), \quad \gamma(p) = m + s\eta(p). \tag{7}$$

It follows that

$$|E|\,\theta = (\mu-m)^2 + 2\sigma(\mu-m)\int_E \xi(p)\,dp - 2s(\mu-m)\int_E \eta(p)\,dp + \int_E \{\sigma\xi(p) - s\eta(p)\}^2\,dp. \tag{8}$$

Consider the particular case $|E| = 1$, that is, E consists of all points in $(0, 1)$ except possibly for a set of measure zero. For instance, we would exclude values of p for which $\xi(p)$ and $\eta(p)$ are not unique. Here

$$\theta = (\mu-m)^2 + \sigma^2 - 2\sigma s I + s^2,$$

where

$$I = \int_0^1 \xi(p)\,\eta(p)\,dp.$$

Now if $G(x)$ differs from $F(x)$ in mean and variance at most, $I = 1$, and

$$\theta = (\mu-m)^2 + (\sigma-s)^2.$$

This is reduced to zero by choosing $m = \mu$ and $s = \sigma$.

In all other cases $I < 1$. To minimize θ we should choose $m = \mu$ and $s = I\sigma$, when θ takes the value $\sigma^2(1-I^2)$. If more than the mean and variance can be chosen we should make our choice to maximize I.

By this criterion the method of fitting $G(x)$ by moments is correct in that it sets $m = \mu$, but is not optimum in taking $s = \sigma$. The method gives $\theta = 2\sigma^2(1-I)$. The optimum procedure is to calculate I and then to choose s less than σ accordingly. However, the method of moments will not be far from optimum if I is near one, i.e. if the standardized forms of $F(x)$ and $G(x)$ are not too different. Suppose $I = 1-\epsilon$, where ϵ is small. By the optimum procedure $\theta = 2\sigma^2\epsilon - \sigma^2\epsilon^2$, and by the method of moments θ is only slightly greater: $\theta = 2\sigma^2\epsilon$.

It appears that ϵ is often small, and in this sense the method of fitting by moments is nearly optimum. I cannot usually be calculated in the cases for which we require approximations. But in fitting a normal distribution to $\sqrt{(2\chi^2)}$, where the χ^2-distribution has 30 d.f., we obtain $I = 0{\cdot}995$. In fitting a normal distribution even to R/R, $\alpha_1 = 0{\cdot}1$, $\alpha_2 = 0{\cdot}05$, we obtain I as large as $0{\cdot}92$.

The more general case, $|E|<1$, will not be discussed here except to point out that the procedure of minimizing θ is very similar. If we write (8) as

$$|E|\theta=(\mu-m)^2+2(\mu-m)(U\sigma-Vs)+\sigma^2X-2\sigma sY+s^2Z,$$

we see we must choose

$$\left.\begin{aligned} m&=\mu+(VY-UZ)\sigma/(V^2-Z),\\ s&=(UV-Y)\sigma/(V^2-Z),\end{aligned}\right\} \tag{9}$$

provided $G(x)$ is non-singular.

If, for example, $F(x)$, $G(x)$ and E are symmetrical about their means, (9) reduces to

$$m=\mu \quad \text{and} \quad s=Y\sigma/Z.$$

5·4. Although the preceding section may be regarded as a refinement and in part a justification of the method of fitting by moments, it cannot be applied as it stands to the lognormal approximation to q. For the lognormal, and for a wide class of distributions, the standard form $\eta(p)$ of § 5·3 does not exist. It exists only if the family $G(x)$ is linear. The product and quotient of standard forms are not a linear family, nor can they be transformed into a linear family. This is evident from the fact that such a product, say $(a+bx)(c+dy)$, or quotient has three parameters (ac, b/a and d/c), while a linear family has only two.

Although we cannot fit a best lognormal to a general product or quotient of variates, the argument above still shows that the method of fitting by moments to q is nearly optimum. The definitions of $\xi(p)$ and $\eta(p)$ must now follow from (7).

The method of fitting to $\log q$ is nearly optimum in another metric, in which the distance between $F(x)$ and $G(x)$ is

$$\int_0^1[\log\{\phi(p)/\gamma(p)\}]^2\,dp.$$

6. Conclusions

We return to the problem of setting limits to

$$q=(x_1x_2\dots x_j)/(x_{j+1}\dots x_n).$$

There is no single answer to the question, what approximation should we use to calculate good limits? We must use judgement in selecting the known distribution used in the calculation: normal, lognormal, or one of those described in § 2. When the family is chosen, the method of fitting by moments to q is recommended. It is recommended only because the metric in which it is nearly optimum is the more natural; in practice there is little difference between the two approximations. Given Tables 1 and 2, the fit to q is easily calculated; without Table 1, the fit to $\log q$ is simpler.

If the number of component variates is larger than two, the lognormal approximation will give satisfactory results when the coefficients of variation of the components are small and not too different. It is therefore a suitable approximation for general use. If the coefficient of variation of q is sufficiently small, the normal approximation gives very similar results.

To calculate the first two moments of q it is necessary to know the coefficient of variation, or quotient of half-range and mean, of each component, and then to combine the values given in Table 2. The percentage points are then given in Table 1, interpolation and extrapolation for other percentage points being discussed in § 5·2. Finally, these points are to be multiplied by

$$(\mu_1\mu_2\dots\mu_j)/(\mu_{j+1}\dots\mu_n).$$

The example we gave in § 1 may now be completed. Suppose, for example, $E = NRN/NN$ (i.e. W is measured with rectangular error, the other components with normal error) and $\alpha_1 = 0{\cdot}02$, $\alpha_2 = 0{\cdot}01$, $\alpha_3 = 0{\cdot}005$, $\alpha_4 = 0{\cdot}015$, $\alpha_5 = 0{\cdot}005$ and $\alpha_6 = 0{\cdot}005$. We find the mean and coefficient of variation of E are 1·000 275 and 2·57 %.

The 1 and 99 % limits to E are 0·941 and 1·061 times its calculated value, i.e. the efficiency has been determined plus or minus about 6 %. The normal approximation gives very similar limits: 0·937 and 1·063.

If, however, α_2 were large in relation to α_1, α_3, etc., we might prefer to use R/N rather than the lognormal as an approximation, and to obtain percentage points from Broadbent (1954).

The author thanks Prof. G. A. Barnard for his encouragment and assistance, Prof. E. S. Pearson for his help in preparing the work for publication, and the Director General of the British Coal Utilization Research Association for permission to publish it.

REFERENCES

AROIAN, L. A. (1947). The probability function of the product of two normally distributed variables. *Ann. Math. Statist.* **18**, 265.

BROADBENT, S. R. (1954). The quotient of a rectangular or triangular and a general variate. *Biometrika*, **41**, 330.

BRUNT, D. (1931). *The Combination of Observations*, 2nd ed. Cambridge University Press.

CRAIG, C. C. (1936). On the frequency function of xy. *Ann. Math. Statist.* **7**, 1.

CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. AktuarTidskr.* **11**, 13.

CRAMÉR, H. (1951). *Mathematical Methods of Statistics*. Princeton University Press.

CREASY, M. A. (1954). Limits for the ratio of means. *J. R. Statist. Soc.* B, **16**, 186.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631.

DERKSEN, J. B. D. (1939). On some infinite series introduced by Tschuprow. *Ann. Math. Statist.* **10**, 380.

FIELLER, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, **24**, 428.

FINNEY, D. J. (1941). On the distribution of a variate whose logarithm is normally distributed. *J. R. Statist. Soc. Suppl.* **7**, 155.

GADDUM, J. H. (1945). Log-normal distributions. *Nature, Lond.*, **156**, 463.

GEARY, R. C. (1930). The frequency distribution of the quotient of two normal variates. *J. R. Statist. Soc.* A, **93**, 442.

HALDANE, J. B. S. (1942). Moments of the distributions of powers and products of normal variates. *Biometrika*, **32**, 226.

JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149.

KOLMOGOROV, A. (1933). Sulla determinazione empirica di una legge di distributione. *G. Ist. ital. Attuari*, **4**, 83.

PEARSON, K. (1922). *Tables of the Incomplete Gamma Function*. Cambridge University Press.

SHELLARD, G. D. (1952). Estimating the product of several random variables. *J. Amer. Statist. Ass.* **47**, 216.

SMIRNOV, N. (1936). Sur la distribution de ω^2. *C.R. Acad. Sci., Paris*, **202**, 449.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung*. Vienna: Deuticke.

WICKSELL, S. D. (1917). On the genetic theory of frequency. *Ark. Mat. Astr. Fys.* **12**, no. 20.

WICKSELL, S. D. (1921). An exact formula for spurious correlation. *Metron*, **1**, 33.

A REJECTION CRITERION BASED UPON THE RANGE†

By C. I. BLISS, W. G. COCHRAN and J. W. TUKEY‡

The experimenter is occasionally faced with an inexplicable, 'aberrant' observation in an otherwise valid set of data. If it is defective and he accepts it—or if it is sound and he rejects it—his results will be biased. By following an objective rule rather than a subjective impression, he can control, and perhaps minimize, his risk of making a wrong decision. The rule proposed here is intended for data consisting of equal-sized sets of replicate measurements, the several sets possibly varying in their means but all being samples from populations with the same variance. It was developed originally to meet the need for a simple rejection criterion for use with several of the bioassays in the *U.S. Pharmacopeia XV* (1955). Other applications could be cited.

For the test which we propose, the range is computed from each set of n measurements, there being k sets in all. The largest range is divided by the sum of all the ranges. The resulting ratio T is compared with the tabular value for the appropriate n and k at the probability level of $P = 0{\cdot}05$. If the observed ratio exceeds the tabular value (cf. Tables 1 and 2), the set represented by this largest range is assumed to contain an aberrant observation or outlier, which is identified by inspection and rejected. Thus the proposed criterion is closely related to Cochran's test (1941) for the largest variance in a series. The application of this rule controls the probability of biasing a result by failing to reject an outlier.

The test may be illustrated numerically with data from two biological assays that were submitted in collaborative studies sponsored by the U.S. Pharmacopeia. The first is an assay of corticotropin from the concentration of ascorbic acid in the adrenal glands of hypophysectomized rats. Seven rats were assigned at random to each of six dosage groups. Each group, in turn, was injected with one of three dosages of the standard preparation (S_1 to S_3) and of the test or unknown preparation (U_1 to U_3). The response y was determined separately from the adrenal glands of each rat. The total for each treatment group is given below, together with its range and sum of squares ($6s^2$):

Dose	S_1	S_2	S_3	U_1	U_2	U_3
Total response Σy	28·92	24·87	20·62	26·95	25·17	20·22
Range ($n = 7$)	0·86	0·79	0·37	1·49	0·80	0·76
Variance × 6, $6s^2$	0·442	0·362	0·117	1·642	0·455	0·467

The range for U_1 (1·49) is the largest and most suspect; when it is divided by the sum of the ranges, we obtain $T = 1{\cdot}49/5{\cdot}07 = 0{\cdot}294$. This exceeds the upper 5 % point of $T = 0{\cdot}288$ for $k = 6$ and $n = 7$, indicating an outlier. The identification of this group as unusual by our proposed range criterion may be checked against Cochran's test; its critical ratio of

† Prepared, in part, in connexion with research sponsored by the Office of Naval Research.

‡ From The Connecticut Agricultural Experiment Station and Yale University, Johns Hopkins University, and Princeton University, respectively.

1·642/3·485 = 0·471 exceeds the upper 1 % point (0·461) in the table by Eisenhart, Hastay & Wallis (1947). Among the individual y's in this group (2·68, 3·90, 4·00, 4·02, 4·06, 4·12 and 4·17), the smallest response not only falls considerably short of the others, but also contributes to a suspiciously small total response for its dose, when compared with the totals for the other groups.

Our second example is from a turbidimetric assay of vitamin B_{12}. Four sets of triplicate tubes were prepared for each of six dosage levels of the reference standard. One set was placed in each of four tube racks, together with triplicate tubes for four dosage levels of a sample or unknown in two of the four racks. Within each rack the tubes were intermingled at random. After incubation overnight, the percentage transmittance of each tube was read in a photometer. The ranges of the $k = 32$ sets of $n = 3$ tubes had the following distribution:

Observed range of 3	0	1	2	3	4	5	6	7	8	13
No. of sets	1	5	6	8	4	2	3	1	1	1

From the total of the ranges (= 113), the observed ratio $T = 13/113 = 0{\cdot}115$. Since $k > 10$, we enter Table 2 with $(k+2)\,T = 34 \times 0{\cdot}115 = 3{\cdot}91$, which much exceeds the value of 3·08, interpolated between 3·03 and 3·11 in the column for $n = 3$. On the basis of all 32 ranges, we attribute the range of 13 to an outlier. The three readings comprising this set show transmittances of 24, 34 and 37 %, from which the reading of 24 % would be rejected as aberrant. The test ratio may now be recomputed with the largest of the 31 remaining ranges in the numerator to obtain $T = 8/100 = 0{\cdot}08$. Since $33 \times 0{\cdot}08 = 2{\cdot}64$ is less than the interpolated value (3·07) for $n = 3$ and $k = 31$, no other observation would be rejected from the series.

Distribution

The critical values of the ratio T in Table 1 have been computed for $k = 2$ to 10 ranges, each determined from a set of $n = 2$ to 10 measurements. Table 2 gives critical values of $(k+2)\,T$ for 10 to 50 ranges. Together they should cover the situations that occur most frequently. The original variates are assumed to be distributed normally with a common standard deviation within groups. Since our criterion is homogeneous, we may take $\sigma = 1$ without loss of generality.

Let $w_1, w_2, \ldots, w_k$ be a set of k ranges from normal samples, each of size n and independently and identically distributed. Let w_* be the largest w and W the sum of the w's. Then

$$T_k = \frac{w_*}{W} = \frac{w_*}{(W - w_*) + w_*}$$

$$= \frac{w_*/(W - w_*)}{1 + w_*/(W - w_*)} = \frac{S_{k-1}}{1 + S_{k-1}},$$

where $$S_{k-1} = \frac{w_*}{w_1 + w_2 + \ldots + w_k - w_*} = \max\left\{\frac{w_i}{w_1 + \ldots + w_{i-1} + w_{i+1} + \ldots + w_k}\right\}.$$

Thus the distribution of T_k is determined by that of S_{k-1}, which is the largest of k identically distributed quantities of the form

$$R_{k-1} = \frac{w_1}{w_2 + w_3 + \ldots + w_k}.$$

The argument used in developing Cochran's test (1941) for the largest variance can be applied to relate the percentage points of S_{k-1} to those of R_{k-1}. The value of S_{k-1} for $P = 0{\cdot}05$ falls between the values of R_{k-1} for

$$P_1 = 0{\cdot}05/k \quad \text{and} \quad P_2 = 1-(0{\cdot}95)^{1/k}.$$

Roughly, P_2 is $0{\cdot}98P_1$ for all k; for $k = 2$, the two values are $P_1 = 0{\cdot}025$ and $P_2 = 0{\cdot}02532$, and for $k = 20$, they are $P_1 = 0{\cdot}0025$ and $P_2 = 0{\cdot}002558$. Since investigation by one of us suggested that the desired value of R_{k-1} was closer to P_1, we have used the compromise probability $P^* = \frac{1}{3}(2P_1+P_2)$, except where $T_k \geqslant \frac{1}{2}$ (that is, $S_{k-1} \geqslant 1$) when P_1 is exact.

The calculation of critical levels of T_k for $P = 0{\cdot}05$ is thus reduced to the calculation of critical values of R_{k-1} for $P = P^*$, from which

$$T_{k,0{\cdot}05} = \frac{R_{k-1,P^*}}{1+R_{k-1,P^*}}.$$

Since the numerator and denominator of R_{k-1} are independent, several approximations are available.

Table 1. *Upper 5 % points of T, the ratio of the largest of k independent normal ranges, each of n observations, to the total of these ranges*

No. of ranges k	No. of observations n in each range								
	2	3	4	5	6	7	8	9	10
2	0·962	0·862	0·803	0·764	0·736	0·717	0·702	0·691	0·682
3	·813	·667	·601	·563	·539	·521	·507	·498	·489
4	·681	·538	·479	·446	·425	·410	·398	·389	·382
5	·581	·451	·398	·369	·351	·338	·328	·320	·314
6	0·508	0·389	0·342	0·316	0·300	0·288	0·280	0·273	0·267
7	·451	·342	·300	·278	·263	·253	·245	·239	·234
8	·407	·305	·267	·248	·234	·225	·218	·213	·208
9	·369	·276	·241	·224	·211	·203	·197	·192	·188
10	·339	·253	·220	·204	·193	·185	·179	·174	·172

Table 2. *Upper 5 % points of $(k+2)\,T$, where T is the ratio of the largest of k independent normal ranges, each of n observations, to the total of these ranges*

No. of ranges k	No. of observations n in each range								
	2	3	4	5	6	7	8	9	10
10	4·06	3·04	2·65	$2{\cdot}44_5$	$2{\cdot}30_5$	2·21	$2{\cdot}14_5$	2·09	2·05
12	4·06	3·03	$2{\cdot}63_5$	$2{\cdot}42_5$	2·29	2·20	2·13	$2{\cdot}07_5$	2·04
15	4·06	3·02	$2{\cdot}62_5$	$2{\cdot}41_5$	2·28	$2{\cdot}18_5$	2·12	$2{\cdot}06_5$	$2{\cdot}02_5$
20	4·13	3·03	2·62	$2{\cdot}41_5$	2·28	$2{\cdot}18_5$	2·11	2·05	2·01
50	4·26	3·11	2·67	2·44	2·29	2·19	2·11	2·06	2·01

APPROXIMATE P^* VALUES OF R_{k-1}

The critical values of R_{k-1} were approximated by two methods discussed below. Most of the values in Table 1 were obtained by method A, and most of those in Table 2 by method B. The two agree moderately well, and intermediate tabular values give some weight to both methods.†

A range or total range is often approximated by a multiple of the square root of a χ^2-variate, i.e. by a χ-variate. Equivalent degrees of freedom ν and scale factors c for a single range are given by Thomson (1953), and for mean ranges by David (1951). From these, the approximate distributions of w_1 and of $\overline{w}$ are

$$w_1 \sim c_1 \chi_1 / \sqrt{\nu_1} \quad \text{and} \quad \overline{w} = \frac{w_2 + w_3 + \ldots + w_k}{k-1} \sim c_2 \chi_2 / \sqrt{\nu_2}.$$

When required, more accurate scale factors for mean or total ranges can be found from $c^2 = d_n^2 + V_n/k$ (David, 1951).

Method A consists of the straightforward approximation

$$R_{k-1} = \frac{w_1}{w_2 + w_3 + \ldots + w_k} = \frac{w_1 \sqrt{\nu_2}}{(k-1)\, c_2 \chi_2} \sim \frac{c_1}{(k-1)\, c_2} \sqrt{F},$$

where $\sqrt{F} = \dfrac{\chi_1}{\chi_2} \sqrt{\dfrac{\nu_2}{\nu_1}}$ and F has ν_1 and ν_2 degrees of freedom. The required values of $\sqrt{F}$ for critical levels of P^* are obtained by interpolation in a table of F.

Method B is less direct and more refined, the χ approximation appearing only in the denominator of R_{k-1}. Let

$$u = \frac{w_2 + w_3 + \ldots + w_k}{(k-1)\, c_2}.$$

Then the expected mean square of u is σ^2, and we can regard

$$R_{k-1} = \frac{w_1}{w_2 + w_3 + \ldots + w_k} = \frac{w_1}{(k-1)\, c_2 u}$$

as the result of studentizing $w_1/(k-1)\, c_2$ with the scale estimator u. We know the result of studentizing a χ-variate (at probability P^*) with the scale-estimator based on ν_2 degrees of freedom. In that case $(F_{\nu_1, \infty})^{\frac{1}{2}}$ is converted into $(F_{\nu_1, \nu_2})^{\frac{1}{2}}$ and the P^* value is increased in the ratio $\sqrt{(F/F_\infty)}$, where F is the P^* value of F_{ν_1, ν_2} and F_∞ is the P^* value of $F_{\nu_1, \infty}$. If we take this same studentizing factor as applicable to the studentization of $w_1/(k-1)\, c_2$ at P^* by u, we have, as the method B approximation,

$$R_{k-1} = \frac{w}{(k-1)\, c_2} \sqrt{\frac{F}{F_\infty}},$$

where w is now the range of n unit normally distributed items at P^* (Pearson, 1942). To improve the accuracy of linear interpolation for non-tabular values of ν_1, ν_2 and P^*, selected values of $\sqrt{(F/F_\infty)}$ were first computed from the five-figure tables of Merrington & Thompson (1943), supplemented by the Hald tables (1952) for larger values of ν_1 and for $P = 0{\cdot}001$.

The two approximations differ by the factor

$$\frac{w/c_1}{\sqrt{F_\infty}} = \frac{w/c_1}{\chi_1/\sqrt{\nu_1}},$$

† It is not claimed that the entries in the tables are correct to the last figure given, but we believe that most of them will not differ from the true values by more than a few units in the last place.

which would be unity if the χ-variate approximation to the numerator range were perfect. Method B has the advantage of allowing for idiosyncrasies in the numerator range, rather than hiding them in a χ-variate approximation. For $k > 2$, the total range in the denominator is approximated more closely by the χ-variate than is the simple range in the numerator.†

The case of $n = 2$

For ranges of 2, $w_1 = \sqrt{2}\,|x|$, where x is a unit normal deviate. Methods A and B are then identical, since w_1 is exactly a χ-variate. The desired critical values are related directly to those of Lord's (1947) statistic

$$u(k-1, 2) = \frac{d_2\,|x|}{\overline{w}} = \frac{(k-1)\,d_2\,|x|}{w_2 + w_3 + \ldots + w_k},$$

where $\overline{w}$ is the mean of $k-1$ ranges, and $d_2 = 1{\cdot}1284$ is the average range of a sample of $n = 2$. The critical values of R_{k-1} are thus $\sqrt{2}/d_2(k-1)$ times those of $u(k-1, 2)$. Despite rather substantial interpolation for P^* at some values of k, we have used Lord's tables in computing the values for $n = 2$ in Tables 1 and 2.

REFERENCES

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen., Lond.*, **11**, 47.

David, H. A. (1951). Further applications of range to the analysis of variance. *Biometrika*, **38**, 393.

Eisenhart, C., Hastay, M. W. & Wallis, W. A. (1947). *Selected Techniques of Statistical Analysis.* New York: McGraw-Hill Book Co.

Hald, A. (1952). *Statistical Tables and Formulas.* New York: John Wiley and Sons.

Lord, E. (1947). The use of range in place of standard deviation in the t-test. *Biometrika*, **34**, 41.

Merrington, M. & Thompson, C. M. (1943). Tables of percentage points of the inverted beta (F) distribution. *Biometrika*, **33**, 73.

Pearson, E. S. (1942). The probability distribution of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301.

Thomson, G. M. (1953). Scale factors and degrees of freedom for small sample sizes for χ-approximation to the range. *Biometrika*, **40**, 449.

Tukey, J. W. (1956). Every man his own studentizer. *Memorandum Report* 58. Statistical Res. Grp. Princeton University.

United States Pharmacopeia XV (1955). Easton, Pa.: Mack Publishing Co.

† Since the completion of the present calculations method B has been studied further by one of the authors (Tukey, 1956). This investigation indicates that, as used here, method B should give slight, but only slight, overestimates of the critical values.

CONFIDENCE INTERVALS FOR A PROPORTION

By EDWIN L. CROW*

U.S. Naval Ordnance Test Station, China Lake, California

1. Summary and definitions

Tables of confidence intervals for a proportion based on the sample proportion are presented, calculated by a slight modification of the method proposed by Sterne (1954), for fixed sample sizes up to 30 and confidence coefficients of 0·90, 0·95 and 0·99. This system is compared, especially in shortness, with Sterne's system, Clopper & Pearson's (1934), and another, intermediate system. It is assumed that a random sample of fixed size n is drawn from an infinite population containing a proportion π of individuals with a given characteristic, that r individuals are observed to have the characteristic, and that it is desired to estimate π by means of a two-sided confidence interval.

While relatively tedious to calculate, Sterne's system and its modification have in common the advantage that no system is possible which has shorter total length of the $n+1$ confidence intervals for $r = 0, 1, \ldots, n$. However, they are at a disadvantage if one-sided confidence intervals and tests are also of interest.

Neyman's definition of a confidence interval $\delta(r)$ (or system of confidence intervals), modified for the case of a single, discrete variable r and one parameter π, consists of the requirement that, whatever be π, the random interval $\delta(r)$ cover the true value of π with a probability at least equal to a prescribed number called the confidence coefficient, say $1-\epsilon$. Neyman showed that construction of a system $\delta(r)$ is equivalent to the determination, for each π, of regions of acceptance $A(\pi)$ such that:

(i) $P\{r \in A(\pi) \mid \pi\} \geqslant 1-\epsilon$.

(ii) Every r is included in at least one $A(\pi)$.

(iii) The set of values of π whose regions $A(\pi)$ contain r is a closed interval.

This interval is the confidence interval for π to be used when the value r is observed.

In the present case of binomial sampling r takes only the integral values $0, 1, 2, \ldots, n$, and the $A(\pi)$ will be taken as successive integers r, $r_1 \leqslant r \leqslant r_2$ say, such that

$$\sum_{r=r_1}^{r_2} p_{n,r}(\pi) \geqslant 1-\epsilon, \quad p_{n,r}(\pi) = \frac{n!}{r!\,(n-r)!}\pi^r(1-\pi)^{n-r}. \tag{1}$$

The end-points r_1 and r_2 are not uniquely determined by (1). Four different further restrictions which do render them unique are considered:

(1) The Clopper–Pearson system, $\delta_1(r)$ say, is determined by choosing 'central' acceptance regions $A_1(\pi)$, so that r_1 is the largest r and r_2 the smallest r with each tail probability not more than $\frac{1}{2}\epsilon$.

(2) The system $\delta_2(r)$ is determined by choosing acceptance regions $A_2(\pi)$ such that

$$r_1 = 0, \quad \sum_{r=r_2+1}^{n} p_{n,r}(\pi) \leqslant \epsilon \quad (0 \leqslant \pi < \pi_{U0}), \tag{2}$$

$$A_2(\pi) \equiv A_1(\pi) \quad (\pi_{U0} \leqslant \pi \leqslant \tfrac{1}{2}), \tag{3}$$

* Now at Boulder Laboratories of the National Bureau of Standards, Boulder, Colorado

where π_{U0} is the Clopper–Pearson upper confidence limit for $r=0$, determined from $(1-\pi_{U0})^n = \frac{1}{2}\epsilon$, and by choosing $A_2(\pi)$ for $\pi > \frac{1}{2}$ as the symmetrical regions consisting of the integers $n-r'$, where r' is in $A_2(\pi')$, $\pi' = 1-\pi$.

(3) The Sterne system $\delta_3(r)$ is determined by choosing acceptance regions $A_3(\pi)$ as those values of r with the largest probabilities of occurring, i.e. the r's are chosen in order, starting with the most probable and continuing in both directions from it until (1) is satisfied. If two values of r have equal probabilities and both cannot be excluded from the acceptance region, then both are included. (The latter provision yields a larger acceptance region than necessary at a finite number of values of π, but it makes the confidence intervals closed without increasing their total length.)

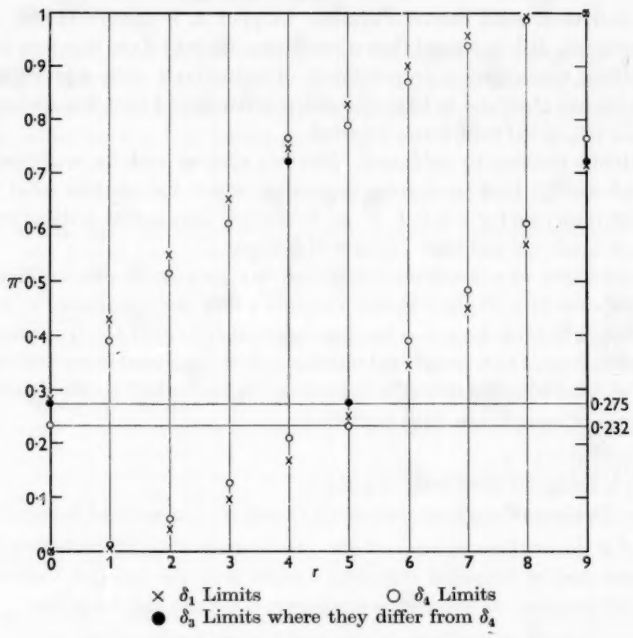


Fig. 1. Confidence intervals $\delta_1, \delta_3, \delta_4$ for $n=9$, $\epsilon=0{\cdot}10$.

(4) The modified Sterne system $\delta_4(r)$ is the same as $\delta_3(r)$ except for the limits determined by the operation called substitution in § 2. We show in § 3 that the Sterne system has the least total length because $A_3(\pi)$ is minimum for each π, except for a set of zero measure. But this minimum property still holds if the substitution of one value of r for another occurs at the *smallest* value of π ($< \frac{1}{2}$) consistent with (i) rather than at the value of π where the two r's become equally probable, subject to the restriction that no r's other than the two of the substitution become involved. For example, in the case illustrated in Fig. 1 (to be discussed later) substitution of $r = 0$ for $r = 5$ may occur as early as $\pi = 0{\cdot}232$ rather than at the value $\pi = 0{\cdot}275$, where $p_{9,0}(\pi) = p_{9,5}(\pi)$. Since a substitution at $\pi < \frac{1}{2}$ simultaneously determines an upper confidence limit for a small value of r and a lower limit for a value nearer $\frac{1}{2}n$, the earlier substitution has the advantage of transferring a given subinterval to a larger r

where it is relatively less important; this is the sole advantage of δ_4 and δ_3. Thus in Fig. 1 the subinterval (0·232, 0·275) is transferred from $r = 0$ to $r = 5$, so that the 90 % confidence interval δ_4 for $r = 0$ is (0, 0·232) and that for $r = 5$ is (0·232, 0·790). However, the natural irregularity of δ_3 is augmented sufficiently by the above definition of δ_4 to cause the 90 % interval to be longer than the corresponding 95 % interval in three independent cases for $n \leqslant 30$ ($n = 7$, $r = 5$; $n = 10$, $r = 5$; $n = 10$, $r = 7$), in one of which ($n = 10$, $r = 5$) the latter interval is covered. (No 95 % interval for $n \leqslant 30$ is longer than the 99 % interval.) Such interesting and theoretically permissible, but practically undesirable, results are prevented in δ_4 and Table 1 by the additional restriction that no substitution is made before the point at which the length of an interval would equal that of a corresponding interval with larger commonly used confidence coefficient.

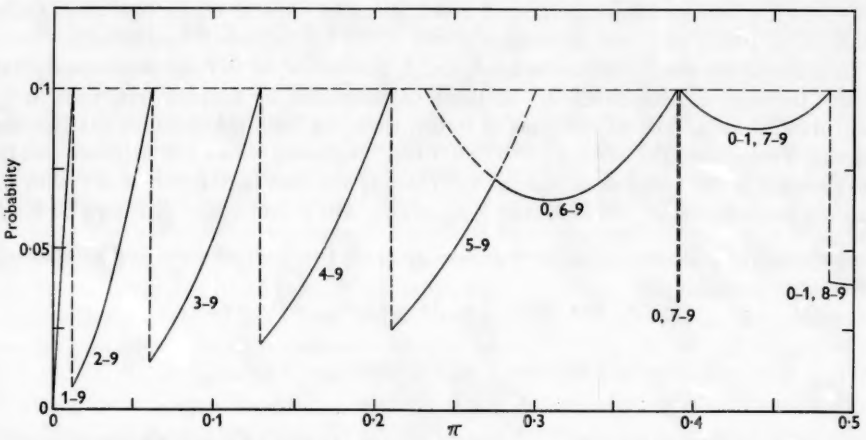


Fig. 2. Probability that the Sterne intervals δ_3 fail to cover the true π ($n = 9, \epsilon = 0{\cdot}10$).

The four systems are defined above by direct satisfaction of condition (i) on acceptance regions. It can be readily shown that conditions (ii) and (iii) are also satisfied by δ_1 and δ_2. However, condition (iii) is not always satisfied by δ_3 or δ_4 for a reason to be discussed near the end of § 2. At least five such cases occur in δ_3 and δ_4 for $n \leqslant 30$: $\epsilon = 0{\cdot}10$, $n = 20$, $r = 0$, and $r = 1$; $\epsilon = 0{\cdot}01$, $n = 27$, $r = 2$; $\epsilon = 0{\cdot}10$, $n = 28$, $r = 2$; $\epsilon = 0{\cdot}01$, $n = 28$, $r = 0$. In the first case, for example, the δ_3 confidence set for $r = 0$ consists of the two intervals (0, 0·127) and (0·141, 0·147). In order to obtain confidence sets which are intervals in such cases also without increasing the size of the acceptance regions, the definition of $A_j(\pi)$ is altered by replacing its least probable value of r by the next less probable value if this can be done without violating condition (i) on $A_j(\pi)$. This has been possible in calculations of δ_3 and δ_4 for $n \leqslant 30$; e.g. in the above case the δ_3 and δ_4 intervals for $r = 0$ are (0, 0·127) and (0, 0·126) respectively by virtue of a confidence interval for $r = 6$ of (0·141, 0·500) rather than (0·147, 0·500).

A similar definition of a confidence interval can be used for the parameter of other one-parameter discrete distributions. A table of δ_4 for the parameter of the Poisson distribution is being constructed.

2. Calculation of tables

Because of the asymptotic normality of the binomial distribution, the differences among the four systems defined above are of interest only for small sample sizes. For the purpose of comparison all four systems were calculated to at least three decimal places for $n \leqslant 20$ (and δ_4 for $n \leqslant 30$), although several tables provide δ_1 in this range as well as beyond (e.g. Hald, 1952). The values of π for which the tail sums of the binomial distribution exactly equal a specified probability are fixed percentage points of the Incomplete Beta-Function given by Thompson (1941) (repeated by Pearson & Hartley, 1954) and Clark (1953). They can also be obtained, accurate to about three decimal places, up to $n = 150$ by linear interpolation in the tables prepared by the National Bureau of Standards (1949) and the U.S. Army Ordnance Corps (1952). By symmetry the acceptance regions need be determined only for $\pi \leqslant \frac{1}{2}$. The confidence limits for δ_1 and δ_2 are given respectively by $100\epsilon/2$ and a combination of $100\epsilon/2$ and 100ϵ percentage points of the Incomplete Beta-Function.

In contrast, the confidence limits for δ_3 and δ_4 correspond to no fixed percentage points of the Incomplete Beta-Function and must be calculated by frequent reference to the individual terms as well as the sums of terms; these are both tabulated in the National Bureau of Standards *Tables* for $\pi = 0{\cdot}01\,(0{\cdot}01)\,0{\cdot}50$. Beginning with $\pi = 0$, we determine the *complements* of the acceptance regions $A_3(\pi)$ (or $A_4(\pi)$), that is, the values of r with the smallest probabilities $p_{n,r}(\pi)$ such that $\sum_r p_{n,r}(\pi) \leqslant \epsilon$. For π sufficiently near zero all such r are consecutive integers ending in n. Hence we enter the table of sums and successively determine:

(1) π_{L1} such that $\sum_{r=1}^{n} p_{n,r}(\pi) \leqslant \epsilon$ in $\pi_{L0} = 0 < \pi \leqslant \pi_{L1}$,

(2) π_{L2} such that $\sum_{r=2}^{n} p_{n,r}(\pi) \leqslant \epsilon$ in $\pi_{L1} < \pi \leqslant \pi_{L2}$,

.....,

up to the largest such value, say $\pi_{L,k-1}$, less than the π_{U0} at which $p_{n,0}(\pi)$ becomes sufficiently small to be included in the sums $\sum_r p_{n,r}(\pi)$. The operation thus performed of transferring one value of r from the complement of $A_3(\pi)$ to $A_3(\pi)$ (no other value of r being involved) is called *elimination*. The point $r = 0$ enters the complement of $A_3(\pi)$ by *addition* at the $\pi = \pi_{U0}$ such that

$$p_{n,0}(\pi) + \sum_{r=k}^{n} p_{n,r}(\pi) = \epsilon \tag{4}$$

if any such π exists, or if not, by *substitution* for $r = k$ at the $\pi = \pi_{U0} = \pi_{Lk}$ such that $p_{n,0}(\pi) = p_{n,k}(\pi)$. The calculation for δ_4 differs only in that the substitution of $r = 0$ for $r = k$ occurs as soon after $\pi = \pi_{L,k-1}$ as

$$p_{n,0}(\pi) + \sum_{r=k+1}^{n} p_{n,r}(\pi) \tag{5}$$

decreases to ϵ.

The calculations of δ_3 and δ_4 are illustrated in Figs. 1 and 2 for $n = 9$, $\epsilon = 0{\cdot}10$; Fig. 1 also shows the corresponding Clopper–Pearson intervals δ_1. There are eliminations at π values of 0·012, 0·061, 0·129, 0·210, 0·390 and 0·485 (lower confidence limits for the eliminated values of r), a substitution at 0·275 (where $p_{9,0} = p_{9,5}$) for δ_3 or 0·232 for δ_4, and an addition at 0·391. In Fig. 2 each curve is labelled with the values of r whose probability sum is plotted. As indicated by dashed extensions of the probability curves, the substitution could be made, with the same minimum length of acceptance region, as late as 0·301 with $p_{9,5}$ in

the sum or as early as 0·232 with $p_{9,0}$ in the sum. As shown in Fig. 1 the substitution simultaneously determines π_{L5} as the lower confidence limit for $r=5$ and its equal, π_{U0}, as the upper confidence limit for $r=0$.

The calculation of δ_3 or δ_4 may proceed from any value of π by determining: (1) the next value of π, if any, at which addition is possible; (2) if not (as is usually the case), the next value of π, if any, at which substitution is possible; and (3) if neither is possible, the next value of π at which elimination is necessary. This procedure assures that the shortest possible acceptance region is maintained for all values of π. For example, if π_{U0} is determined by substitution, we determine next whether any π satisfies

$$p_{n,0}(\pi)+p_{n,1}(\pi)+\sum_{r=k+1}^{n} p_{n,r}(\pi)=\epsilon. \tag{6}$$

If so, the solution is π_{U1} by addition. If not, we determine whether $p_{n,1}(\pi)$ decreases to $p_{n,k+1}(\pi)$ before (5), which has a minimum, increases to ϵ, and in the affirmative determine $\pi_{U1}=\pi_{L,k+1}$ for δ_4 by substitution where (5) with $p_{n,k+1}(\pi)$ replaced by $p_{n,1}(\pi)$ decreases to ϵ. In the negative $\pi_{L,k+1}$ is determined where (5) increases to ϵ. We continue in this way up to $\pi=\frac{1}{2}$, determining whether an addition, substitution, or elimination occurs next. The δ_3 or δ_4 confidence interval for π when r 'successes' are observed is then (π_{Lr}, π_{Ur}).

The reason for the occurrence in the initial definition of δ_3 or δ_4 of confidence sets containing two intervals may now be discussed. The probability sums such as in Fig. 2 that include two tails always have a smooth minimum which occurs at a value of π different from that at which its largest term (from the left tail) becomes equal to the next largest term (from the right tail). All calculations support the conjecture that the latter value of π is always larger ($\pi<\frac{1}{2}$). Then ϵ may be just such a size that the two-sided sum increases from its minimum through ϵ where its largest term is still from the left tail and hence should be eliminated rather than a term on the right. This would give two intervals of π with $A(\pi)$ containing the corresponding left-tail value of r. For each two-sided sum there would always be such an ϵ, but for any particular ϵ it would occur only rarely and can evidently be avoided as specified in § 1.

Table 1, comprising the δ_4 system of confidence intervals for $n = 1\,(1)\,30$ and confidence coefficients 0·90, 0·95 and 0·99, was calculated in the above manner using 2-point to 7-point Lagrangian interpolation in the National Bureau of Standards *Tables*. In addition, Robert S. Gardner checked the entire calculation independently on a high-speed electronic calculator using automatic coding. The total time with the latter was about 50 hr. spread over two or three weeks, while desk calculation required about 200 hr. spread over many months. All differences (mainly of 1 in the third decimal place due to inadequacy of linear interpolation) were reconciled, so that Table 1 should have no error. The tables of Thompson and Clark were used for checking the limits that arise with all the probability ϵ in one tail.

3. COMPARISON OF SYSTEMS OF CONFIDENCE INTERVALS

The system δ_2 is shorter than δ_1 under any conceivable definition of 'shortness' because all its intervals are contained in those of δ_1 and some of δ_1 contain points outside the corresponding intervals of δ_2. On the other hand, though the δ_1 and δ_2 intervals more often than not contain the corresponding δ_3 intervals, there are fourteen intervals out of the total of 690 δ_3 intervals calculated for $n\leqslant 20$ which contain the corresponding intervals of both δ_1 and δ_2, and 18 (63) are longer than the corresponding intervals $\delta_1(\delta_2)$.

Table 1. *Table of confidence limits for a proportion*

Calculated by Edwin L. Crow, Eleanor G. Crow and Robert S. Gardner according to a modification of a proposal of Theodore E. Sterne.

n	r	Confidence coefficient (%) 90	95	99
1	0	0·000	0·000	0·000
	1	·100	·050	·010
2	0	0·000	0·000	0·000
	1	·051	·025⁺	·005⁺
	2	·316	·224	·100
3	0	0·000	0·000	0·000
	1	·035⁻	·017	·003
	2	·196	·135⁺	·059
	3	·464	·368	·215⁺
4	0	0·000	0·000	0·000
	1	·026	·013	·003
	2	·143	·098	·042
	3	·320	·249	·141
	4	·500	·473	·316
5	0	0·000	0·000	0·000
	1	·021	·010	·002
	2	·112	·076	·033
	3	·247	·189	·106
	4	·379	·343	·222
	5	·621	·500	·398
6	0	0·000	0·000	0·000
	1	·017	·009	·002
	2	·093	·063	·027
	3	·201	·153	·085⁻
	4	·333	·271	·173
	5	·458	·402	·294
	6	·655⁺	·598	·464
7	0	0·000	0·000	0·000
	1	·015⁻	·007	·001
	2	·079	·053	·023
	3	·170	·129	·071
	4	·279	·225⁺	·142
	5	·316	·341	·236
	6	·500	·446	·357
	7	·684	·623	·500
8	0	0·000	0·000	0·000
	1	·013	·006	·001
	2	·069	·046	·020
	3	·147	·111	·061
	4	·240	·193	·121
	5	·255⁻	·289	·198
	6	·418	·315⁺	·293
	7	·582	·500	·410
	8	·745⁺	·685⁻	·549
9	0	0·000	0·000	0·000
	1	·012	·006	·001
	2	·061	·041	·017
	3	·129	·098	·053
	4	·210	·169	·105⁺
	5	·232	·251	·171
	6	·390	·289	·250
	7	·485⁻	·442	·344
	8	·609	·557	·402
	9	·768	·711	·598
10	0	0·000	0·000	0·000
	1	·010	·005⁺	·001
	2	·055⁻	·037	·016
	3	·116	·087	·048
	4	·188	·150	·093
	5	·222	·222	·150
	6	·341	·267	·218
	7	·352	·381	·297
	8	·500	·397	·376
	9	·648	·603	·488
	10	·778	·733	·624
11	0	0·000	0·000	0·000
	1	·010	·005⁻	·001
	2	·049	·033	·014
	3	·105⁻	·079	·043
	4	·169	·135⁺	·084
	5	·197	·200	·134
	6	·302	·250	·194
	7	·315⁺	·333	·262
11	8	0·423	0·369	0·340
	9	·577	·500	·407
	10	·685⁻	·631	·500
	11	·803	·750	·641
12	0	0·000	0·000	0·000
	1	·009	·004	·001
	2	·045⁺	·030	·013
	3	·096	·072	·039
	4	·154	·123	·076
	5	·184	·181	·121
	6	·271	·236	·175⁻
	7	·294	·294	·235⁻
	8	·398	·346	·302
	9	·500	·450	·321
	10	·602	·550	·445⁺
	11	·706	·654	·555⁻
	12	·816	·764	·679
13	0	0·000	0·000	0·000
	1	·008	·004	·001
	2	·042	·028	·012
	3	·088	·066	·036
	4	·142	·113	·069
	5	·173	·166	·111
	6	·246	·224	·159
	7	·276	·260	·213
	8	·379	·327	·273
	9	·455⁺	·413	·302
	10	·530	·480	·406
	11	·621	·566	·477
	12	·724	·673	·571
	13	·827	·775⁻	·698
14	0	0·000	0·000	0·000
	1	·007	·004	·001
	2	·039	·026	·011
	3	·081	·061	·033
	4	·131	·104	·064
	5	·163	·153	·102
	6	·224	·206	·146
	7	·261	·206	·195⁻

The observed proportion in a random sample of size n is r/n. The table gives the lower confidence limit for the population proportion π, as a function of n and r.

The upper confidence limit = 1 − (lower confidence limit, entered with $n-r$ instead of r).

Table 1 (*cont.*)

n	r	Confidence coefficient (%) 90	95	99
14	8	0·355⁻	0·312	0·249
	9	·406	·371	·286
	10	·422	·389	·364
	11	·578	·500	·392
	12	·635⁻	·611	·500
	13	·739	·688	·608
	14	·837	·794	·714
15	0	0·000	0·000	0·000
	1	·007	·003	·001
	2	·036	·024	·010
	3	·076	·057	·031
	4	·122	·097	·059
	5	·154	·142	·094
	6	·205⁺	·191	·135⁻
	7	·247	·191	·179
	8	·325⁺	·294	·229
	9	·325⁺	·332	·273
	10	·400	·369	·328
	11	·500	·448	·373
	12	·600	·552	·461
	13	·675⁻	·631	·539
	14	·753	·698	·627
	15	·846	·809	·727
16	0	0·000	0·000	0·000
	1	·007	·003	·001
	2	·034	·023	·010
	3	·071	·053	·029
	4	·114	·090	·055⁺
	5	·147	·132	·088
	6	·189	·178	·125⁺
	7	·235⁺	·178	·166
	8	·299	·272	·212
	9	·305⁺	·272	·261
	10	·381	·352	·295⁺
	11	·450	·429	·357
	12	·550	·500	·421
	13	·619	·571	·475⁺
	14	·695⁻	·648	·549
	15	·765⁻	·728	·643
	16	·853	·822	·736
17	0	0·000	0·000	0·000
	1	·006	·003	·001
	2	·032	·021	·009
	3	·067	·050	·027
	4	·107	·085⁻	·052
	5	·140	·124	·082
	6	·175⁺	·166	·117
	7	·225⁺	·166	·155⁺
	8	·277	·253	·197
	9	·290	·254	·242
	10	·364	·337	·242
	11	·432	·406	·338
	12	·500	·456	·380
	13	·568	·511	·413
	14	·636	·583	·500
	15	·710	·663	·587
	16	·775⁻	·746	·654
	17	·860	·834	·758
18	0	0·000	0·000	0·000
	1	·006	·003	·001
	2	·030	·020	·008
	3	·063	·047	·025⁺
	4	·101	·080	·049
	5	·135⁻	·116	·077
	6	·163	·156	·110
	7	·216	·157	·145⁺
	8	·257	·236	·184
	9	·277	·242	·226
	10	·349	·325⁻	·228
	11	·416	·375⁻	·314
	12	·464	·381	·318
	13	·518	·444	·397
	14	·581	·556	·466
	15	·651	·619	·534
	16	·723	·675⁺	·603
	17	·784	·758	·682
	18	·865⁺	·843	·772
19	0	0·000	0·000	0·000
	1	·006	·003	·001
	2	·028	·019	·008
	3	·059	·044	·024
	4	·095⁺	·075⁺	·046
19	5	0·130	0·110	0·073
	6	·151	·147	·103
	7	·209	·150	·137
	8	·238	·222	·173
	9	·265⁺	·232	·212
	10	·337	·312	·218
	11	·386	·345⁻	·293
	12	·386	·365⁻	·305⁺
	13	·440	·426	·383
	14	·560	·500	·436
	15	·614	·574	·485⁻
	16	·663	·635⁺	·545⁻
	17	·735⁻	·684	·617
	18	·791	·768	·695⁻
	19	·870	·850	·782
20	0	0·000	0·000	0·000
	1	·005⁺	·003	·001
	2	·027	·018	·008
	3	·056	·042	·023
	4	·090	·071	·044
	5	·126	·104	·069
	6	·141	·140	·098
	7	·201	·143	·129
	8	·221	·209	·163
	9	·255⁻	·222	·200
	10	·325⁻	·293	·209
	11	·358	·293	·274
	12	·367	·351	·293
	13	·422	·411	·363
	14	·500	·467	·399
	15	·578	·533	·424
	16	·633	·589	·500
	17	·672	·649	·576
	18	·745⁺	·707	·625⁺
	19	·797	·778	·707
	20	·874	·857	·791
21	0	0·000	0·000	0·000
	1	·005⁺	·002	·000
	2	·026	·017	·007
	3	·054	·040	·022
	4	·086	·068	·041
	5	·121	·099	·065⁺

The observed proportion in a random sample of size n is r/n. The table gives the lower confidence limit for the population proportion π, as a function of n and r.

The upper confidence limit = 1 − (lower confidence limit, entered with $n-r$ instead of r).

Table 1 (*cont.*)

n	r	Confidence coefficient (%) 90	95	99
21	6	0·130	0·132	0·092
	7	·191	·137	·122
	8	·191	·197	·155⁻
	9	·245⁻	·213	·189
	10	·306	·276	·201
	11	·306	·276	·257
	12	·353	·338	·283
	13	·407	·398	·339
	14	·458	·449	·347
	15	·542	·494	·409
	16	·593	·545⁻	·466
	17	·647	·602	·534
	18	·694	·662	·591
	19	·755⁺	·724	·653
	20	·809	·787	·717
	21	·877	·863	·799
22	0	0·000	0·000	0·000
	1	·005⁻	·002	·000
	2	·024	·016	·007
	3	·051	·038	·021
	4	·082	·065⁻	·039
	5	·115⁻	·094	·062
	6	·115⁻	·126	·088
	7	·181	·132	·116
	8	·181	·187	·147
	9	·236	·205⁺	·179
	10	·289	·260	·194
	11	·289	·264	·242
	12	·340	·326	·273
	13	·393	·383	·318
	14	·444	·418	·334
	15	·500	·424	·396
	16	·556	·500	·450
	17	·607	·576	·495⁺
	18	·660	·611	·546
	19	·711	·674	·604
	20	·764	·736	·666
	21	·819	·795⁻	·727
	22	·885⁺	·868	·806
23	0	0·000	0·000	0·000
	1	·005⁻	·002	·000
	2	·023	·016	·007
	3	·049	·037	·020
	4	·078	·062	·038
23	5	0·110	0·090	0·059
	6	·110	·120	·084
	7	·173	·127	·111
	8	·173	·178	·140
	9	·228	·198	·171
	10	·273	·247	·187
	11	·274	·255⁻	·229
	12	·328	·317	·265⁺
	13	·381	·360	·298
	14	·431	·360	·323
	15	·478	·409	·384
	16	·521	·457	·420
	17	·569	·543	·429
	18	·619	·591	·500
	19	·672	·640	·571
	20	·726	·683	·614
	21	·772	·745⁺	·677
	22	·827	·802	·735⁻
	23	·890	·873	·813
24	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·022	·015⁺	·006
	3	·047	·035⁻	·019
	4	·075⁻	·059	·036
	5	·105⁻	·086	·057
	6	·105⁻	·115⁻	·080
	7	·165⁻	·122	·106
	8	·165⁻	·169	·133
	9	·221	·191	·163
	10	·259	·234	·181
	11	·264	·246	·216
	12	·317	·308	·257
	13	·370	·339	·280
	14	·413	·347	·313
	15	·447	·396	·362
	16	·447	·443	·364
	17	·553	·500	·416
	18	·577	·557	·464
	19	·630	·604	·536
	20	·683	·653	·584
	21	·736	·692	·636
	22	·779	·754	·687
	23	·835⁺	·809	·741
	24	·895⁺	·878	·819
25	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·021	·014	·006
	3	·045⁻	·034	·018
	4	·072	·057	·034
	5	·101	·082	·054
	6	·101	·110	·077
	7	·158	·118	·101
	8	·158	·161	·127
	9	·214	·185⁺	·155⁺
	10	·246	·222	·175⁺
	11	·255⁻	·238	·205⁺
	12	·307	·296	·245⁺
	13	·360	·317	·245⁺
	14	·389	·336	·305⁻
	15	·389	·384	·342
	16	·432	·431	·352
	17	·500	·475⁻	·403
	18	·568	·525⁺	·451
	19	·611	·569	·500
	20	·638	·616	·549
	21	·693	·664	·597
	22	·745⁺	·697	·648
	23	·786	·762	·695⁺
	24	·842	·815⁻	·755⁻
	25	·899	·882	·825⁻
26	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·021	·014	·006
	3	·043	·032	·017
	4	·069	·054	·033
	5	·097	·079	·052
	6	·097	·106	·073
	7	·151	·114	·097
	8	·151	·154	·122
	9	·209	·180	·149
	10	·233	·212	·170
	11	·247	·230	·195⁻
	12	·299	·282	·234
	13	·342	·282	·234
	14	·342	·325⁺	·298
	15	·377	·374	·322
	16	·419	·421	·342
	17	·460	·458	·393
	18	·540	·494	·438
	19	·581	·535⁻	·474

The observed proportion in a random sample of size n is r/n. The table gives the lower confidence limit for the population proportion π, as a function of n and r.

The upper confidence limit = 1 − (lower confidence limit, entered with $n-r$ instead of r).

Table 1 (*cont.*)

n	r	Confidence coefficient (%) 90	95	99
26	20	0·623	0·579	0·513
	21	·658	·626	·558
	22	·701	·675⁻	·607
	23	·753	·718	·658
	24	·791	·770	·702
	25	·849	·820	·766
	26	·903	·886	·830
27	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·020	·013	·006
	3	·042	·031	·017
	4	·066	·052	·032
	5	·093	·076	·050
	6	·093	·101	·070
	7	·145⁺	·110	·093
	8	·145⁺	·148	·117
	9	·204	·175⁻	·143
	10	·221	·202	·166
	11	·239	·223	·185⁻
	12	·291	·269	·224
	13	·326	·269	·225⁻
	14	·326	·316	·284
	15	·365⁺	·364	·298
	16	·407	·402	·332
	17	·447	·430	·383
	18	·500	·437	·413
	19	·553	·500	·419
	20	·593	·563	·461
	21	·635⁻	·585⁺	·539
	22	·674	·636	·581
	23	·709	·684	·616
	24	·761	·731	·668
	25	·796	·777	·703
	26	·855⁻	·825⁺	·775⁺
	27	·907	·890	·834
28	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·019	·013	·005⁺
	3	·040	·030	·016
	4	·064	·050	·031
	5	·089	·073	·048
	6	·089	·098	·068
	7	·139	·106	·089
	8	·139	·142	·112
	9	·197	·170	·137
	10	·208	·192	·162
	11	·232	·217	·175⁺
	12	·284	·258	·214
	13	·310	·259	·218
	14	·312	·307	·272
	15	·355⁻	·355⁻	·272
	16	·396	·381	·323
	17	·435⁺	·384	·364
	18	·473	·424	·364
	19	·527	·463	·408
	20	·565⁻	·537	·449
	21	·604	·576	·500
	22	·645⁺	·616	·551
	23	·688	·643	·592
	24	·716	·693	·636
	25	·768	·741	·677
	26	·799	·783	·728
	27	·861	·830	·782
	28	·911	·894	·838
29	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·018	·012	·005⁺
	3	·039	·029	·015⁺
	4	·062	·049	·030
	5	·086	·070	·046
	6	·086	·094	·065⁺
	7	·134	·103	·086
	8	·134	·136	·108
	9	·189	·166	·132
	10	·189	·184	·157
	11	·225⁻	·211	·165⁺
	12	·276	·247	·206
	13	·294	·251	·211
	14	·303	·299	·260
	15	·345⁻	·339	·263
	16	·385⁺	·339	·316
	17	·425⁻	·374	·346
	18	·463	·413	·354
	19	·500	·451	·397
	20	·537	·500	·438
	21	·575⁺	·549	·477
	22	·615⁻	·587	·523
	23	·655⁺	·626	·562
	24	·697	·661	·603
	25	·721	·701	·646
	26	·775⁺	·749	·684
	27	·811	·789	·737
	28	·866	·834	·789
	29	·914	·897	·840
30	0	0·000	0·000	0·000
	1	·004	·002	·000
	2	·018	·012	·005⁺
	3	·037	·028	·015⁻
	4	·059	·047	·028
	5	·083	·068	·045⁻
	6	·083	·091	·063
	7	·129	·100	·083
	8	·129	·131	·104
	9	·182	·163	·127
	10	·182	·175⁺	·151
	11	·219	·205⁺	·151
	12	·265⁻	·236	·198
	13	·265⁻	·244	·206
	14	·295⁻	·292	·249
	15	·336	·324	·256
	16	·376	·324	·308
	17	·416	·364	·329
	18	·446	·403	·345⁻
	19	·476	·440	·388
	20	·508	·476	·430
	21	·545⁻	·524	·462
	22	·584	·560	·495⁻
	23	·624	·597	·531
	24	·664	·636	·570
	25	·705⁺	·676	·612
	26	·735⁺	·708	·655⁺
	27	·781	·756	·690
	28	·818	·795⁻	·744
	29	·871	·837	·794
	30	·917	·900	·849

The observed proportion in a random sample of size n is r/n. The table gives the lower confidence limit for the population proportion π, as a function of n and r.

The upper confidence limit = 1 − (lower confidence limit, entered with $n-r$ instead of r).

Neyman's (1937) definition of the shortest system and the short unbiased system of confidence intervals, and Scheffé's (1942*a*) added definition of the shortest unbiased system, deal with such less clear-cut comparisons in the case of continuous variables with probability densities. However, these definitions have not been applied to discrete variables except indirectly, by changing the given variable to a continuous one by adding an auxiliary variable with a rectangular distribution, as reviewed by Pearson (1950). In particular it can be easily shown that none of the four systems considered in this paper is either shortest or unbiased. While it is true that shortest systems generally do not exist even for continuous variables, Eudey (1949) obtained shortest unbiased systems for π by the above device. Tables of these intervals do not appear to be available.

The confidence interval systems δ_3 and δ_4 have an optimum shortness property in a geometric sense. If an acceptance region $A(\pi)$ of any system is considered to have length equal to $r_2 - r_1 + 1$, then as π varies from 0 to 1, the acceptance regions sweep out a region in the (r, π) plane composed of not more than $2n+1$ rectangles which may be called the confidence belt. Thus in Fig. 1 the rectangles containing $A_4(\pi)$, 17 in all, would be, reading from bottom to top:

$$-0{\cdot}5 \leqslant r \leqslant 0{\cdot}5,\ 0 \leqslant \pi < 0{\cdot}012; \quad -0{\cdot}5 \leqslant r \leqslant 1{\cdot}5,\ 0{\cdot}012 \leqslant \pi < 0{\cdot}061; \quad \ldots;$$

$$-0{\cdot}5 \leqslant r \leqslant 4{\cdot}5,\ 0{\cdot}210 \leqslant \pi \leqslant 0{\cdot}232; \quad 0{\cdot}5 \leqslant r \leqslant 5{\cdot}5,\ 0{\cdot}232 \leqslant \pi < 0{\cdot}390;$$

$$0{\cdot}5 \leqslant r \leqslant 6{\cdot}5,\ 0{\cdot}390 \leqslant \pi \leqslant 0{\cdot}391; \quad 1{\cdot}5 \leqslant r \leqslant 6{\cdot}5,\ 0{\cdot}391 < \pi < 0{\cdot}485;$$

$$1{\cdot}5 \leqslant r \leqslant 7{\cdot}5,\ 0{\cdot}485 \leqslant \pi \leqslant 0{\cdot}515; \quad \ldots; \quad 8{\cdot}5 \leqslant r \leqslant 9{\cdot}5,\ 0{\cdot}988 < \pi \leqslant 1{\cdot}000.$$

If the lower and upper confidence limits are denoted by π_{Lr} and $\pi_{Ur} = 1 - \pi_{L,n-r}$, then the area of the confidence belt is

$$n + 1 - 2(1.\pi_{L1} + 1.\pi_{L2} + \ldots + 1.\pi_{Ln}) = \sum_{r=0}^{n} (\pi_{Ur} - \pi_{Lr}),$$

that is, the total length of the $n+1$ confidence intervals. Now $A_3(\pi)$ and $A_4(\pi)$ are chosen so that their length is a minimum whatever π (except for a set of measure zero), for a given probability $1-\epsilon$ is attained with as few terms as possible by including the largest possible terms. Since the area of the confidence belt is the integral of the length of $A(\pi)$ over the fixed interval $(0, 1)$, it is as small as possible with $A_3(\pi)$ or $A_4(\pi)$. Thus *among all possible systems of confidence intervals, no system has shorter total length than the systems δ_3 and δ_4.*

The total length of the system δ_4 in Fig. 1 is

$$2(0{\cdot}232 + 0{\cdot}379 + 0{\cdot}454 + 0{\cdot}481 + 0{\cdot}558) = 4{\cdot}208.$$

Its 'mean length' $\bar{d}_4$ (all r's equally weighted) is 0·4208, while the mean length $\bar{d}_1$ of δ_1 is 0·4705. The mean lengths $\bar{d}_i$ are tabulated on p. 433 as functions of sample size for $\epsilon = 0{\cdot}10$. The differences $\bar{d}_1 - \bar{d}_2$ are positive and decrease monotonely after a maximum at $n=3$. The differences $\bar{d}_2 - \bar{d}_3$ are positive beyond $n=4$ and oscillate, larger values being associated with 'additions' in the calculation of δ_3. The differences $\bar{d}_1 - \bar{d}_3$, likewise oscillate, with an absolute maximum of 0·087 at $n=3$ and an absolute minimum in this range of 0·015 at $n=26$. The percentage reduction in mean length from δ_1 to δ_3 or δ_4 varies from a minimum of 5% at $n=1$ to a maximum of 12% at $n=3$, 5 and 6 and is 6% at $n=30$. The differences $\bar{d}_1 - \bar{d}_3$ for $\epsilon = 0{\cdot}05$ and $\epsilon = 0{\cdot}01$ are essentially the same as for $\epsilon = 0{\cdot}10$ beyond $n=6$; the percentage reductions are then less than for $\epsilon = 0{\cdot}10$ but decrease more slowly (to 4% at $n=30$ for $\epsilon = 0{\cdot}01$). The large-sample confidence intervals based on the normal approximation to the

binomial have, for $n=30$ and $\epsilon=0{\cdot}10$, a mean length that is 0·013 less than $\bar{d}_3$ and 0·029 less than $\bar{d}_1$.

n	1	2	3	4	5	6	7	8
$\bar{d}_1$	0·950	0·834	0·740	0·667	0·612	0·565	0·528	0·497
$\bar{d}_3=\bar{d}_4$	·900	·755	·653	·604	·540	·498	·490	·451
$\bar{d}_1-\bar{d}_2$	·050	·079	·087	·063	·060	·042	·033	·030
$\bar{d}_2-\bar{d}_3$	·000	·000	·000	·000	·012	·025	·005	·016

n	10	12	14	16	18	20	25	30
$\bar{d}_1$	0·448	0·410	0·380	0·355	0·335	0·317	0·283	0·257
$\bar{d}_3=\bar{d}_4$	·417	·373	·350	·328	·306	·296	·267	·242
$\bar{d}_1-\bar{d}_2$	·023	·017	·013	·010	·008	·005	·004	·003
$\bar{d}_2-\bar{d}_3$	·008	·020	·017	·017	·021	·016	·012	·012

The mean lengths $\bar{d}_i$ may be compared by graphical integration with the like mean length of Eudey's shortest unbiased system for the case he gives explicitly (Graph 13), $n=5$, $1-\epsilon=0{\cdot}80$. The mean lengths of δ_3 and Eudey's system are the same to graphical accuracy, 0·437; this is 7 % less than $\bar{d}_2$ and 17 % less than $\bar{d}_1$.

The advantage of δ_4 over δ_3 can be measured for given ϵ and n by comparing the averages over all r from 1 to $[\frac{1}{2}n]$ of the ratio of the length $d_{i,r}$ of the confidence interval for a given r divided by r/n, that is,

$$\bar{R}_i = \sum_{r=1}^{[\frac{1}{2}n]} \frac{d_{i,r}}{r},$$

where $[\frac{1}{2}n]$ denotes the largest integer contained in $\frac{1}{2}n$. A few values are indicated below for $\epsilon=0{\cdot}10$:

n	5	10	15	20	25	30
$\bar{R}_1$	2·54	2·09	1·94	1·69	1·59	1·45
$\bar{R}_1-\bar{R}_3$	0·29	0·11	0·12	0·12	0·08	0·07
$\bar{R}_3-\bar{R}_4$	−0·05	0·04	0·07	0·03	0·06	0·06

The negative value of $\bar{R}_3-\bar{R}_4$ results from the necessary omission of the term for $r=0$ in the averaging.

Confidence intervals are often chosen by minimizing their expected length irrespective of the parameter value (Kendall, 1946, p. 72; Scheffé, 1942*b*, 1943). A disadvantage of this criterion is its dependence on the particular form of the parameter, e.g. π or $2\arcsin\sqrt{\pi}$. Scheffé (1942*b*) showed that when it is applied to the logarithm of the ratio of variances from two normal populations, it gives the shortest unbiased system. The expected lengths of δ_1, δ_2, δ_3, δ_4, and Eudey's shortest unbiased system, say δ_5, for $n=5$, $\epsilon=0{\cdot}20$ were evaluated using the National Bureau of Standards *Tables* and are graphed in Fig. 3 as a function of π.

In this example at least, δ_3 and δ_4 are uniformly better than δ_2 according to this criterion, and δ_2 is better than δ_1 except for $\pi = 0$ and 1, but the shortest unbiased system δ_5 is not uniformly best. Of the four systems based on the discrete variable, δ_4 approximates δ_5 best with respect to the desirable property of relatively short expected length for π near 0 and 1.

The Sterne system δ_3 (or δ_4) minimizes the expected length of confidence interval if π has an *a priori* distribution which makes all values of r equally probable. Such an *a priori* distribution is the uniform distribution over (0, 1).

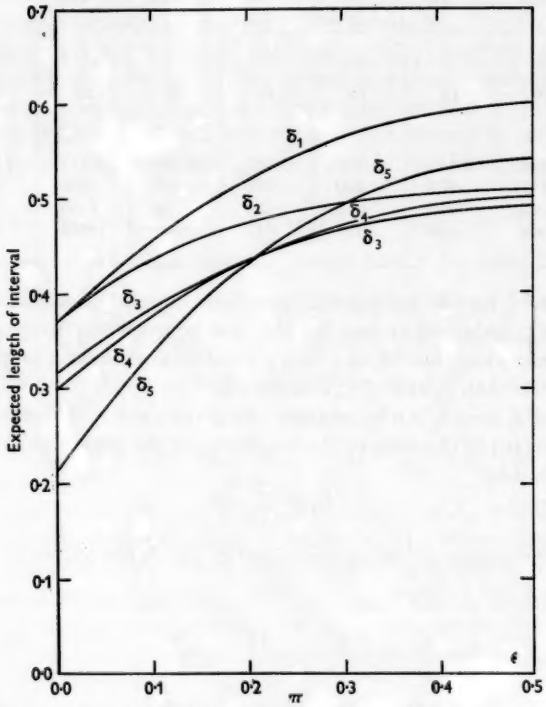


Fig. 3. Expected lengths of various confidence intervals for $n = 5$, $1 - \epsilon = 0{\cdot}80$.

The systems may be compared on the basis of the probability of covering the true parameter value π, or of failing to cover, as shown in Fig. 2 for δ_3 and δ_4, $n = 9$, $\epsilon = 0{\cdot}10$. The probabilities for δ_1 and δ_2 either fall below the curves shown or coincide with them. The probability for δ_4 rises to ϵ at only 16 points in (0, 1). The probability for δ_1 never rises to ϵ, having an absolute maximum of 0·088. If a uniform *a priori* distribution of π over (0, 1) is assumed, the mean probabilities of covering π by δ_1, δ_2, δ_3 and δ_4 are 0·965, 0·950, 0·938 and 0·935 respectively.

I should like to acknowledge the collaboration of Eleanor G. Crow and Robert S. Gardner in the calculations and to thank Theodore E. Sterne for communicating his results to me prior to publication.

REFERENCES

CLARK, R. E. (1953). Percentage points of the Incomplete Beta Function. *J. Amer. Statist. Ass.* **48**, 831.

CLOPPER, C. J. & PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404.

EUDEY, M. W. (1949). On the treatment of discontinuous random variables. Technical Report No. 13, Statistical Laboratory, University of California, Berkeley.

HALD, A. (1952). *Statistical Tables and Formulas*. New York: John Wiley and Sons.

KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, **2**. London: Charles Griffin and Company.

NATIONAL BUREAU OF STANDARDS (1949). *Tables of the Binomial Probability Distribution*. Washington: U.S. Government Printing Office.

NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans.* A, **236**, 333.

PEARSON, E. S. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika*, **37**, 383.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**, Table 16. Cambridge University Press.

SCHEFFÉ, H. (1942*a*). On the theory of testing composite hypotheses with one constraint. *Ann. Math. Statist.* **13**, 280.

SCHEFFÉ, H. (1942*b*). On the ratio of the variances of two normal populations. *Ann. Math. Statist.* **13**, 371.

SCHEFFÉ, H. (1943). On solutions of the Behrens-Fisher problem based on the *t*-distribution. *Ann. Math. Statist.* **14**, 35.

STERNE, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**, 275.

THOMPSON, C. M. (1941). Tables of percentage points of the Incomplete Beta-Function. *Biometrika*, **32**, 151.

U.S. ARMY ORDNANCE CORPS (1952). *Tables of the Cumulative Binomial Probabilities*. Ordnance Corps Pamphlet ORDP 20-1. Washington 25, D.C.: Office of Technical Services, Department of Commerce (Order No. PB111389).

SERIAL CORRELATION IN REGRESSION ANALYSIS. II

BY G. S. WATSON* AND E. J. HANNAN
Australian National University, Canberra, A.C.T.

1. INTRODUCTION

In a previous paper (Watson, 1955)† a theoretical analysis was made of the effect of making a wrong presumption concerning the correlation matrix of the residuals from a regression of the form

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{k1} \\ \vdots & & \vdots \\ x_{1N} & \dots & x_{kN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}, \tag{1·1}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

when the vector $\mathbf{u}$ has zero expectation and

$$E(\mathbf{uu}') = \sigma^2\boldsymbol{\alpha} \quad (\boldsymbol{\alpha} \text{ non-singular}).$$

In particular, bounds were obtained for the following quantities:

(*a*) The efficiency of the estimates of the β_i (see equation (1·3·5)). This is measured by the ratio of the determinantal values of the covariance matrices of the best linear unbiased estimates and the actual estimates used.

(*b*) The significance points for the test of the hypothesis

$$\mathbf{f}_i'\boldsymbol{\beta} = \phi_i \quad (i = 1, \dots, h \leqslant k),$$

where $\mathbf{f}_i$, ϕ_i are given and the $\mathbf{f}_i$ are linearly independent. The test statistic used was that appropriate to the case $\boldsymbol{\alpha} = \mathbf{I}$ (see equations (1·4·33), (1·4·37), (1·4·44) and (1·4·54)).

The bounds in each case were shown to depend upon the latent roots (particularly the extreme latent roots) of the matrix $\mathbf{H}\boldsymbol{\alpha}\mathbf{H}'$, where $\mathbf{H}$ is such that $\mathbf{H}\boldsymbol{\gamma}\mathbf{H}' = \mathbf{I}$ and $\boldsymbol{\gamma}$ is the presumed correlation matrix. These latent roots are therefore the latent roots of $\boldsymbol{\alpha}\boldsymbol{\gamma}^{-1}$.

In this paper we go on to study cases where the errors, u_i, are generated by certain stationary processes whose form or parameters are incorrectly prescribed *a priori*. In § 2 the spectral theory of stochastic processes is used to study the latent roots of $\boldsymbol{\alpha}\boldsymbol{\gamma}^{-1}$. In § 3 the functions of these latent roots which are needed to examine the bounds to quantities (*a*) and (*b*) above are evaluated when the true and assumed error processes are autoregressive and moving-average. In § 4 the results of § 3 are discussed and illustrated with tables and graphs. In general, the bounds are wide, indicating that the effect of not knowing the true form of the error process may be severe. In § 5 certain asymptotic results are established to show that, in many cases which occur in practice, the bounds may be attained.

* The results in §§ 3, 4 were obtained while one of us (G.S.W.) was a member of the Department of Applied Economics, Cambridge.

† For convenient reference to equations in Part I of this paper, we merely prefix their numbers by unity; thus equation (1·2·5) is equation (2·5) of Part I.

2. THE LATENT ROOTS OF $\boldsymbol{\alpha\gamma^{-1}}$

Consider the case where the u_t are generated by a stationary process with absolutely continuous spectral function, the spectral density being

$$f(\omega) = \sum_{-\infty}^{\infty} \rho_s e^{is\omega}, \tag{2·1}$$

the ρ_s being the serial correlations of the process.

If we assume that the true spectral density is $g(\omega)$ with

$$g(\omega)^{-1} = k \left| \sum_{0}^{\infty} b_j e^{-ij\omega} \right|^2 \quad (b_0 = 1), \tag{2·2}$$

then we shall apply on autoregressive transformation to (1·1) obtaining a system of errors which will be, at least asymptotically, of the form

$$v_t = k^{\frac{1}{2}} \sum_{0}^{\infty} b_j u_{t-j}. \tag{2·3}$$

The spectral density of the v_t will then be

$$\frac{f(\omega)}{g(\omega)} \frac{\sigma^2}{\sigma_v^2} = h(\omega) \frac{\sigma^2}{\sigma_v^2}, \tag{2·4}$$

where σ_v^2 is the variance of the v_t.

The correlation matrix corresponding to (2·4) will be $\sigma_v^{-2}\sigma^2 \mathbf{H\alpha H'}$, so that the function generating, as Fourier coefficients, the elements of $\mathbf{H\alpha H'}$ is $h(\omega)$. For the class of processes we shall be considering it can be shown that the latent roots of $\mathbf{H\alpha H'}$ (that is, the latent roots of $\boldsymbol{\alpha\gamma^{-1}}$) will be, asymptotically, the values of $h(\omega)$ for equidistant values of the argument (in the range $-\pi$ to π) (see Whittle, 1951; Grenander, 1952).

The extreme latent roots of $\boldsymbol{\alpha\gamma^{-1}}$ can then be approximated by the least and greatest values of $h(\omega)$ in the range $-\pi$ to π, while the sum of the latent roots will be, approximately,

$$H = \frac{N}{2\pi} \int_{-\pi}^{\pi} h(\omega)\, d\omega = N \frac{\sigma_v^2}{\sigma^2}. \tag{2·5}$$

We shall indicate the extreme values of $h(\omega)$ in $-\pi$ to π by h_l and h_u.

In the original form of this paper (Watson, 1951), a different approach was used to obtain the results of this paper. Autoregressive and moving average processes of the first and second order were approximated by processes with commutative covariance matrices whose latent roots and vectors were known. The present approach is more general but gives no guide to the effect of the approximations. The earlier work showed that for the cases to be considered below the use of the extreme values of (2·4) and the integral (2·5) will result in errors which are sufficiently small, for our purposes, even for samples of only 15 or 20. Similarly, in formulae such as (1·2·5), the average of the $N-k$ largest (or smallest) roots of $\boldsymbol{\alpha\gamma^{-1}}$ appear. We shall replace this by the average of all of the latent roots. Again the effect will become small as N increases relative to k.

We shall consider cases where the spectral densities $f(\omega)$ and $g(\omega)$ are rational functions of $\cos\omega$, which corresponds to the cases where we have

$$\sum_{0}^{p} \lambda_j u_{t-j} = \sum_{0}^{q} \mu_j \epsilon_{t-j}, \tag{2·6}$$

the ϵ_i being independent random variates. We shall use λ_j and μ_j for the true process as indicated in (2·6) and l_j, m_j for the corresponding parameters in the assumed process. We shall use p' and q' for the maximum lags in the assumed process.

The spectral density of a process, u_i, of the form (2·6) is (Doob, 1953)

$$f(\omega) = \frac{\left|\sum_0^q \mu_j e^{ij\omega}\right|^2}{\left|\sum_0^p \lambda_j e^{ij\omega}\right|^2} \frac{\sigma_\epsilon^2}{\sigma^2} \quad (\lambda_0 = \mu_0 = 1). \tag{2·7}$$

Here σ_ϵ^2 is the variance of the process ϵ_i. When the assumed process is also of the form (2·6), the evaluation of (2·5) or the extreme values of (2·4) is relatively straightforward, though care has to be exercised in some cases.

The bounds which are to be developed are homogeneous functions of degree zero in the latent roots so that we can neglect the factors k and $\sigma_\epsilon^2 \sigma^{-2}$, which appear in $h(\omega)$, without affecting the result, and this will be done below.

Then (2·5) becomes

$$H = N \frac{\sum_{ij=0}^{\infty} b_i b_j \rho_{|i-j|}}{\sum_{ij=0}^{p} \lambda_i \lambda_j \rho_{|i-j|}} \sum_{i=0}^{q} \mu_i^2. \tag{2·8}$$

3. The effects of a wrong prescription for an error process of the form (2·6)

The bounds to (*a*) and (*b*) (of § 1) obtained in Part I of the present study depended on the extreme roots and the sum of the roots of the matrix $\boldsymbol{\alpha\gamma}^{-1}$ *after the removal of the roots corresponding to any regressor vectors which were latent vectors* of the covariance matrix of the transformed residuals. In using the approximations h_l and h_u for the extremes and H for the sum of the roots we shall, of course, be neglecting the effect of the removal of these roots. We shall also be neglecting the effect of the value of k, in the bounds to (*b*). The effect of resulting errors will, of course, vanish asymptotically, but it also appears to be small for quite small values of N. We shall evaluate the bounds to the efficiency only for the case where one regressor vector is not a latent vector. The more general case, where none of the k regressor vectors are latent vectors, can be treated, approximately, by raising the bounds for the case $k = 1$ to the power k.

Below, for certain processes of the general form (2·3), we give the value of $h(\omega)$, H, h_l and h_u. In addition we show, sometimes for particular values of the parameters involved, the lower bound to the efficiency of this estimate, E_l (the upper bound is, of course, unity), and the factors

$$f_l = N \frac{h_l}{H}, \quad f_u = N \frac{h_u}{H}. \tag{3·1}$$

These last factors, when multiplied by the significance point for t^2 appropriate to the case where $h(\omega)$ is unity and the distribution of the u_i is normal, give approximations to the true significance point (for the chosen level). This approximation was discussed in the last paragraph of Part I of this investigation.

(i) $p = 1, q = 0, p' = 1, q' = 0$.

Here it will be convenient to put $-\lambda_1 = \rho_1$ and $-l_1 = r_1$:

$$h(\omega) = \frac{1+r_1^2-2r_1\cos\omega}{1+\rho_1^2-2\rho_1\cos\omega},$$

$$H = \frac{N}{1-\rho_1^2}(1-2\rho_1 r_1+r_1^2),$$

$$h_u = \left(\frac{1-r_1}{1-\rho_1}\right)^2, \quad h_l = \left(\frac{1+r_1}{1+\rho_1}\right)^2 \quad (\rho_1 \geqslant r_1),$$

$$h_u = \left(\frac{1+r_1}{1+\rho_1}\right)^2, \quad h_l = \left(\frac{1-r_1}{1-\rho_1}\right)^2 \quad (\rho_1 \leqslant r_1),$$

$$E_l = \frac{(1-\rho_1^2)^2(1-r_1^2)^2}{(1+r_1^2-4r_1\rho_1+\rho_1^2+r_1^2\rho_1^2)^2},$$

$$\left.\begin{aligned} f_u &= \left(\frac{1-r_1}{1-\rho_1}\right)^2 \frac{1-\rho_1^2}{1-2r_1\rho_1+r_1^2} \\ f_l &= \left(\frac{1+r_1}{1+\rho_1}\right)^2 \frac{1-\rho_1^2}{1-2r_1\rho_1+r_1^2} \end{aligned}\right\} \quad (\rho_1 \geqslant r_1).$$

When $\rho \leqslant r$ the bounds f_u and f_l are interchanged. The values of E_l, f_u and f_l for various values of ρ_1 and r_1 are shown respectively in Tables 1–3 below.

(ii) $p = 2, q = 0, p' = 1, q' = 0$.

Again it will be convenient to put $-\lambda_1(1+\lambda_2)^{-1} = \rho_1$ and $-l_1 = r_1$:

$$h(\omega) = \frac{1+r_1^2-2r_1\cos\omega}{1+\lambda_1^2+\lambda_2^2+2\lambda_1(1+\lambda_2)\cos\omega+2\lambda_2\cos 2\omega},$$

$$H = N\frac{1+r_1^2-2r_1\rho_1}{1-\lambda_1^2-\lambda_2^2-2\lambda_1\lambda_2\rho_1}.$$

Case (*a*). $r_1 = 0, \left|\dfrac{\lambda_1(1+\lambda_2)}{4\lambda_2}\right| \geqslant 1.$

$$h_u = \frac{1}{(1+\lambda_2)^2(1-\rho_1)^2}, \quad h_l = \frac{1}{(1+\lambda_2)^2(1+\rho_1)^2},$$

$$E_l = (1-\rho_1^2)^2(1+\rho_1^2)^{-2} \quad \text{(see Table 1 for } r_1 = 0).$$

Case (*b*). $r_1 = 0, \left|\dfrac{\lambda_1(1+\lambda_2)}{4\lambda_2}\right| \leqslant 1.$

The extreme values h_l and h_u are then obtained from one of those under case (*a*) and

$$\frac{1}{(1-\lambda_2)^2\left\{1-\rho_1^2\dfrac{(1+\lambda_2)^2}{4\lambda_2}\right\}}$$

(depending on the signs of λ_1 and λ_2).

E_l will not be written down but is shown for some λ_1 and λ_2 in Table 4.

Case (*c*). $r_1 = \rho_1$.

$$h_u = \frac{1}{(1-\lambda_2)^2}, \quad h_l = \frac{1}{(1+\lambda_2)^2} \quad (\lambda_2 \geqslant 0),$$

$$h_u = \frac{1}{(1+\lambda_2)^2}, \quad h_l = \frac{1}{(1-\lambda_2)^2} \quad (\lambda_2 \leqslant 0),$$

$$E_l = \left(\frac{1-\lambda_2^2}{1+\lambda_2^2}\right)^2.$$

See Table 1 for $r_1 = 0$ reading λ_2 for ρ_1.

The effect of these incorrect prescriptions of the error process on the test of significance are discussed in the following section.

(iii) $p = 0$, $q = 2$, $p' = 0$, $q' = 1$.

We shall put

$$\rho_1 = \frac{\mu_1(1+\mu_2)}{1+\mu_1^2+\mu_2^2}, \quad \rho_2 = \frac{\mu_2}{1+\mu_1^2+\mu_2^2}, \quad r_1 = \frac{m_1}{1+m_1^2},$$

$$h(\omega) = \frac{1+\mu_1^2+\mu_2^2+2\mu_1(1+\mu_2)\cos\omega+2\mu_2\cos 2\omega}{1+m_1^2+2m_1\cos\omega}.$$

We consider only the case where $r_1 = \rho_1$. Since $r_1 \leqslant \frac{1}{2}$ this imposes the restriction $\rho_1 \leqslant \frac{1}{2}$:

$$H = \frac{N(1+\mu_1^2+\mu_2^2)}{1+m_1^2}\left\{1+\frac{\rho_2}{\rho_1}\left(\frac{1-2\rho_1^2}{(1-4\rho_1^2)^{\frac{1}{2}}}-1\right)\right\}.$$

The extremes h_u and h_l are obtained from

$$\frac{1+\mu_1^2+\mu_2^2}{1+m_1^2}\left\{1+\frac{2\rho_2}{\rho_1^2}[(1-2\rho_1^2)^{\frac{1}{2}}-1]\right\},$$

and one of

$$\frac{1+\mu_1^2+\mu_2^2}{1+m_1^2}\left(1+\frac{2\rho_2}{1-2\rho_1}\right), \quad (\rho_1\rho_2 \text{ positive}),$$

$$\frac{1+\mu_1^2+\mu_2^2}{1+m_1^2}\left(1+\frac{2\rho_2}{1+2\rho_1}\right), \quad (\rho_1\rho_2 \text{ negative}).$$

Only the bound E_l was evaluated and this is shown in Table 5 and Fig. 1.

(iv) $p = 0$, $q = 1$, $p' = 1$, $q' = 0$.

Again we put $\mu_1(1+\mu_1^2)^{-1} = \rho_1$ and $-l_1 = r_1$:

$$h(\omega) = (1+\mu_1^2+2\mu_1\cos\omega)(1+l_1^2+2l_1\cos\omega),$$

$$H = N\{(1+r_1^2)(1+\mu_1^2)-2r_1\mu_1\}.$$

Case (*a*). $\left|\frac{(\mu_1-r_1)(1-\mu_1 r_1)}{4\mu_1 r_1}\right| \geqslant 1.$

The extremes of $h(\omega)$ are then

$$(1+r_1)^2(1-\mu_1)^2, \quad (1-r_1)^2(1+\mu_1)^2.$$

$$E_l = (1-r_1^2)^2(1-\mu_1^2)^2(1+r_1^2-2r_1\mu_1+\mu_1^2+r_1^2\mu_1^2)^{-2},$$

which is shown in Table 1, putting $\mu_1 = \rho_1$.

Case (*b*). $\left|\frac{(\mu_1-r_1)(1-\mu_1 r_1)}{4\mu_1 r_1}\right| \leqslant 1.$

The extremes of $h(\omega)$ are

$$(1+r_1^2)(1+\mu_1^2)+\frac{(\mu_1-r_1)^2(1-\mu_1 r_1)^2}{4\mu_1 r_1},$$

and one of

$$(1+r_1)^2(1-\mu_1)^2 \quad (r_1 \leqslant \mu_1),$$
$$(1-r_1)^2(1+\mu_1)^2 \quad (r_1 \geqslant \mu_1).$$

The efficiency E_l is shown, for this case, in Table 6.

Other cases may be considered, perhaps for restricted values of the parameters, by suitably identifying their parameters with those of examples (i)–(iv) above. This is because of the inverse relationship of the spectral density functions of the autoregressive and moving-average processes. Some examples are

$$p=1,\ q=1,\ p'=0,\ q'=0 \rightarrow \text{(i)},$$
$$p=0,\ q=1,\ p'=0,\ q'=1 \rightarrow \text{(i)},$$
$$p=1,\ q=1,\ p'=0,\ q'=1 \rightarrow \text{(ii)},$$
$$p=1,\ q=0,\ p'=0,\ q'=1 \rightarrow \text{(ii)},$$
$$p=1,\ q=1,\ p'=1,\ q'=0 \rightarrow \text{(iii)},$$
$$p=0,\ q=2,\ p'=0,\ q'=0 \rightarrow \text{(iv)},$$
$$p=0,\ q=0,\ p'=2,\ q'=0 \rightarrow \text{(iv)}.$$

4. Comments on the results of § 3

From the last paragraph of § 3, Table 1 is relevant to several cases but is labelled for the case where a first-order autoregressive error process, first serial correlation ρ_1, is mistaken for a process of the same form but with first serial correlation r_1.

Table 1 is the most widely applicable of the three tables as it is of use in cases (i), (ii$a$), (ii$c$), (iv$a$), and in other situations which, by the last paragraph of § 3 are equivalent to these cases. The discussion is given in terms of case (i). The only striking feature of Table 1 is the poor performance of the high values of r_1. *Thus it appears that the first difference transformation may lead to very low efficiencies unless* $\rho_1 = 1$; the zero efficiency at $r_1 = 1$ is due to the approximation. It will be noted that, except for high ρ_1 and r_1, a value of r_1 within 0·1 or 0·2 of ρ_1 leads to efficiencies of better than 70%. It is only when ρ_1 is high that the use of a near exact value of r_1 is imperative for a reasonably efficient analysis.

Table 4 covers the case where a second-order autoregressive process is mistaken for a process of independent variates. When the procedure appropriate to independent residuals is used, the first column of Table 1 or Table 4 apply, according as $|(\lambda_1+\lambda_1\lambda_2)\frac{1}{4}\lambda_2^{-1}| \geqslant 1$ or $\leqslant 1$. Table 4 also shows the corresponding values of $(1-\rho_1^2)^2(1+\rho_1^2)^{-2}$, appropriate when $|(\lambda_1+\lambda_1\lambda_2)\frac{1}{4}\lambda_2^{-1}| \geqslant 1$, for comparison. Thus while there are second-order autoregressive processes which give direct least squares an efficiency which is dependent only on the first serial correlation $\rho_1 = -\lambda_1(1+\lambda_2)^{-1}$ and is high when ρ_1 is small, there are others which make the efficiency very low even when ρ_1 is small.

Table 5 covers the case where a second-order moving average is mistaken for a first-order moving average with the correct first serial correlation ρ_1. The results are displayed graphically in Fig. 1. For each pair of values of μ_1 and μ_2 the correlograms of the true and assumed error processes have been drawn. The graphs have been arranged in the same order as in Table 5, and on each the lower bound to the efficiency is given. An examination of Fig. 1 shows that an apparently better agreement between the true and assumed correlograms may lead to a worse lower bound to the efficiency. This suggests that the common practice of comparing correlograms may not be the best procedure.

As has been seen in § 2 the lower bound to the efficiency is simply related to the spectral densities rather than to the correlograms. For the cases which we are considering the lower bound to the efficiency can be zero only if the true spectral density has a zero. This is seen to be so for all the cases with a zero lower bound in Table 5. This lower bound will of course be attained (or nearly attained) for regressor vectors which are latent vectors of the correlation matrix of the transformed residuals corresponding to the zero (or near zero) latent root, but it may also be attained, at least asymptotically, for other regressor vectors also (see Grenander, 1952, p. 568, and § 5 below).

Table 6 covers the case where a first-order moving average is mistaken for a simple Markoff process. We have two cases according as $|(\mu_1 - r_1)(1 - \mu_1 r_1)(4\mu_1 r_1)^{-1}|$ is $\geqslant$ or $\leqslant 1$. The first case is covered by Table 1 with $\mu_1 = \rho_1$ and the second by Table 6. Both show that the efficiency falls off more rapidly as ρ_1 and r_1 diverge, than for cases where the two processes are of the same type, as is to be expected.

Finally, Tables 2 and 3 show the correcting factors which may be applied to the significance points for the test of significance of a regression coefficient (based on t^2) when the true and assumed error processes are first-order autoregressive. The meaning of these tables may be seen from an example. Suppose that the t-test has been made when the error has a first serial correlation of 0·5 but when the estimation procedure has been based on an assumed first serial correlation of 0·7. Since $r_1 = 0{\cdot}7$, $\rho_1 = 0{\cdot}5$, the true significance point lies between $1{\cdot}22t^2$ and $0{\cdot}34t^2$, where t^2 is the tabular value—assuming, of course, that the approximation is exact.

In order to find the order of magnitude of the possible effects of serial correlation on the t-tests and to investigate the range of validity of the approximation, either numerical examples or the asymptotic approach of § 5 must be considered. In the calculations shown in Table 7, only the extreme significance levels of t-tests using the tabular significance points were found (for the case of only one regressor in addition to the mean). This avoids the difficulty of inverse interpolation. Although Table 7 is not extensive, it shows that the first approximation is quite a useful guide. The exact results were found by using the method of Pitman & Robbins (1949), the calculations being performed by the EDSAC in the Mathematical Laboratory, University of Cambridge. The second approximation there shown was obtained by making the first two small-sample moments agree for the case where there was only one regressor apart from the mean. For this case the adjustment corresponding to the use of the f_l and f_u also makes the first two moments correct, *asymptotically*, so that a fair agreement is to be expected. The validity of the analysis corresponding to the assumption that the errors follow a first-order process whose first serial correlation equals that of the true error process, when in fact the true error process is of the second order, will now be examined by considering the disturbances in the significance levels of the t-test of regression coefficients. Let it first be supposed that the true errors follow a second-order autoregressive process with $\lambda_1 = -1{\cdot}3$ and $\lambda_2 = 0{\cdot}3$. This is the process suggested by Orcutt (1948) for Tinbergen's series. Since $\lambda_2 < \frac{1}{4}\lambda_1^2$, the correlogram of this process does not have harmonic oscillations. It will be seen further that it has a weak central tendency. The first serial correlation of the process is unity, so that the lower bound to the efficiency of a least-squares analysis is zero. Here we are supposing that a first-order autoregressive transformation based on $r_1 = 1$ has been used, i.e. that the analysis has been carried out on the first differences of the series involved. In this situation the lower bound to the efficiency may be computed and is 0·70. In a sample of 20, the use of the second approximation for deter-

mining the bias in the significance level of the classical t-test at 5 % shows that the true significance level may be almost as small as 1 % and as great as 16 %. As a second example, we will suppose that the errors follow a second-order autoregressive process with $\lambda_1 = -0{\cdot}5$ and $\lambda_2 = -0{\cdot}4$. The correlogram does not have a harmonic oscillation. If least squares had been used directly, the efficiency lower bound would have been 0·02. However, if the analysis had been carried on the data after the use of a first-order autoregressive transformation with $\rho_1 = -\lambda_1(1+\lambda_2)^{-1} = 0{\cdot}83$, the efficiency lower bound is raised to 0·54. In the latter circumstances, the true significance level of a t-test made with the tabulated 5 % point may range from 0·3 % to slightly greater than 20 %. These calculations confirm the idea that better results are obtained in this case by the use of a first-order autoregressive transformation than without it, but they also show clearly that the performance may still be poor even when the first serial correlation is known free of error.

As a last example, consider an error process which is a first-order moving average with $\mu_1 = 0{\cdot}6$, while it is presumed to be a Markoff process with $r_1 = \mu_1(1+\mu_1^2)^{-1} = 0{\cdot}44$, when 20 observations are available. The ratios of the least and greatest roots to the mean root are 0·330 and 1·49 respectively. Since the 5 % point of t^2 with 18 degrees of freedom is 4·41, the first approximation tells us that a test using this point may have any significance level between $P\left(F_{1,18} \geqslant \frac{4{\cdot}41}{0{\cdot}330}\right)$ and $P\left(F_{1,18} \geqslant \frac{4{\cdot}41}{1{\cdot}49}\right)$, i.e. lie between 0·4 and 10·5 %. These probabilities agree well with 0·3 and 10·5 %, found by using the second approximation above.

5. Regression on analytic functions

Grenander (1954) and Grenander & Rosenblatt (1954) consider regressor vectors $\mathbf{x}_t$ (orthonormal) which are subjected to restrictions which effectively require the existence of the limits,

$$r_{\mu\nu}(h) = \lim_{n\to\infty} \sum_{t=1}^{n-h} x_{\mu(t+h)} x_{\nu t} \quad (\mu, \nu = 1, \ldots, k),$$

and permit the representation

$$\mathbf{R}_h = [r_{\mu\nu}(h)] = \int_{-\pi}^{\pi} e^{ih\theta}\, d\mathbf{M}(\theta),$$

where $\mathbf{M}(\theta)$ is a matrix of functions with $\mathbf{M}(\theta_2) - \mathbf{M}(\theta_1)$ Hermitian non-negative definite for $\theta_2 > \theta_1$. It is presumed that $\mathbf{R}_0$ is non-singular. Introducing $\mathbf{N}(\theta) = \mathbf{M}(\theta+) - \mathbf{M}(-\theta-)$ (which is symmetric and real), it is shown by Grenander & Rosenblatt that if $\mathbf{N}(\theta)$ has only $s \leqslant k$ points of increase $\theta_1, \ldots, \theta_s$ such that

(i) $d\mathbf{N}(\theta_j) = [n_{\mu\nu}(j)] = \mathbf{N}_j$ is non-null,

(ii) $\mathbf{N}_i \mathbf{N}_j = $ null matrix $(i \neq j)$,
$= \mathbf{N}_i \quad (i = j)$,

then the least-squares procedure gives estimates of the regression coefficients which have, asymptotically, the same covariance matrix as best linear unbiased estimates. These conditions are also shown to be necessary for this to be so.

In fact it is easy to show (under fairly weak restriction on $f(\omega)$)* that

$$\lim_{n\to\infty} [\mathbf{x}'_\mu \boldsymbol{\alpha} \mathbf{x}_\nu] = \left[\sum_{j=1}^{s} f(\theta_j)\, n_{\mu\nu}(j)\right] = \sum_{j=1}^{s} f(\theta_j)\, \mathbf{N}_j, \tag{5·1}$$

* In particular, $f(\omega)$ must not have zeros at the points θ_j (see Grenander, 1952, p. 568).

and taking into account the properties of the $\mathbf{N}_j$ it is evident that the latent roots of $\mathbf{X}'\boldsymbol{\alpha}\mathbf{X}$ are, in the limit, the $f(\theta_j)$ repeated k_j times (where $k_j = \operatorname{tr}\mathbf{N}_j$, integral; $\sum_j k_j = k$). The sufficiency part of the theorem then follows immediately from (1·3·1) (see Grenander, 1954, p. 263).

These results may be used to suggest asymptotic approximations to the quantities for which we have above provided bounds. It will be seen below that these bounds are often attained.

An important example of a set of regressors satisfying Grenander & Rosenblatt's theorem are the orthogonal (Legendre) polynomials for which $s = 1$ and $\theta_1 = 0$.

Applying the asymptotic result (5·1) to the formula (1·2·4) for the fractional bias of the ith regression coefficient, it is seen to reduce to

$$\frac{N - \sum_{j=1}^{s} k_j f(\theta_j)}{N-k} - \sum_{j=1}^{s} f(\theta_j)\, n_{ii}(j). \tag{5·2}$$

The statistic R of (1·4·49) for testing all the k regression coefficients, takes the form

$$R = \frac{f(\theta_1)\,\eta_1^2 + \ldots + f(\theta_s)\,\eta_k^2}{\sum_1^N \alpha_i \eta_i^2}, \tag{5·3}$$

with each $f(\theta_j)$ repeated k_j times, since the roots ν_i of $\mathbf{X}'\boldsymbol{\alpha}\mathbf{X}$ (see (1·4·50) and (1·4·52)) are asymptotically the $f(\theta_j)$. Thus T of (1·4·46) is given by $T = R(1-R)^{-1}$, with R expressed by (5·3). The relationship of T with the bounding random variables, given in (1·4·54), depends on the relation of $f(\theta)_1, \ldots, f(\theta_s)$ with $\alpha_1, \ldots, \alpha_N$. The approximations which lead to the factors (3·1) for the bounding random variables suggest that here T should be approximated by

$$\frac{\frac{1}{k}\Sigma k_j f(\theta_j)}{\frac{1}{N-k}(N - \Sigma k_j f(\theta_j))} \frac{\sum_1^k \eta_i^2}{\sum_{k+1}^N \eta_i^2}. \tag{5·4}$$

If the regressors are orthogonal polynomials, (5·2) reduces to

$$\frac{N(1-f(0))}{N-k} \approx 1 - f(0), \tag{5·5}$$

and the constant factor of (5·4) becomes

$$\frac{(N-k)f(0)}{N-kf(0)} \approx f(0). \tag{5·6}$$

Supposing further that the error process is Markoff, parameter ρ and that, as Grenander & Rosenblatt's theorem would suggest, least squares has been used, the value of $f(0)$ by (2·7) is $(1-\rho)^{-2}(1-\rho^2)$. Then (5·5) becomes

$$\frac{-N}{N-k}\frac{2\rho}{1-\rho}. \tag{5·7}$$

Watson (1951) showed that this was the minimum bias for $\rho > 0$ (maximum for $\rho < 0$) for this type of error process. Turning to the statistic T discussed above, it is seen that, for $\rho > 0$,

$f(0)$ is essentially the largest of the roots $\alpha_1, \ldots, \alpha_N$ so that T is approximately the upper bounding random variable of (1·4·54). The approximation of using the factor (5·6), here $(1+\rho)(1-\rho)^{-1}$, is exactly the same as using f_u of case 3(a) with $r = 0$, since least squares is assumed to have been used. When $\rho < 0$, T is distributed like the lower bound of (1·4·54). Thus, in this case, the bounding values are good approximations to the true values because $f(0)$ happens to coincide with f_u or f_l. This also happens when the errors are from a second order autoregressive process, falling in case 3(a), but not when it belongs to case 3(b). Similar observations may be made in other cases.

It is therefore evident that the bounds are attained in cases which must occur frequently in practice—analytic regressors and simple error processes. This is certainly a surprising result. Since these bounds are often wide, it is clear that no single rule can reasonably hope to cover all cases.

Our thanks are due to Mr S. Gill of the Cambridge Mathematical Laboratory for calculating certain Pitman-Robbins series on the EDSAC and to Mrs E. M. Chambers and the Computing Staff of the Department of Applied Economics, University of Cambridge, for the remainder of the calculations.

REFERENCES

DOOB, J. L. (1953). *Stochastic Processes*. New York: John Wiley.
GRENANDER, U. (1952). *Ark. Mat.* **1**, 555.
GRENANDER, U. (1954). *Ann. Math. Statist.* **25**, 252.
GRENANDER, U. & ROSENBLATT, M. (1954). *Proc. Nat. Acad. Sci., Wash.*, **40**, 812.
ORCUTT, G. H. (1948). *J. R. Statist. Soc.* B, **10**, 1.
PITMAN, E. J. G. & ROBBINS, H. E. (1949). *Ann. Math. Statist.* **20**, 552.
WATSON, G. S. (1951). *Serial Correlation in Regression Analysis*. Institute of Statistics, North Carolina. Mimeograph Series, no. 49.
WATSON, G. S. (1955). *Biometrika*, **42**, 327.
WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist and Wiksells.

Table 1. *Lower bound to efficiency of estimates. Case* $p = 1, q = 0, p' = 1, q' = 0$

ρ_1 \ r_1	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	1·00	0·96	0·85	0·70	0·52	0·36	0·22	0·12	0·05	0·01
0·1	0·96	1·00	0·96	0·84	0·68	0·49	0·31	0·19	0·07	0·02
0·2	0·85	0·96	1·00	0·96	0·83	0·64	0·43	0·24	0·11	0·02
0·3	0·70	0·84	0·96	1·00	0·95	0·81	0·58	0·35	0·16	0·04
0·4	0·52	0·68	0·83	0·95	1·00	0·94	0·76	0·50	0·24	0·06
0·5	0·36	0·49	0·64	0·81	0·94	1·00	0·92	0·68	0·36	0·09
0·6	0·22	0·31	0·43	0·58	0·76	0·92	1·00	0·89	0·55	0·16
0·7	0·12	0·17	0·24	0·35	0·50	0·68	0·89	1·00	0·81	0·30
0·8	0·05	0·07	0·11	0·16	0·24	0·36	0·55	0·81	1·00	0·60
0·9	0·01	0·02	0·02	0·04	0·06	0·09	0·16	0·30	0·60	1·00

Table 2. $f_u = \left(\frac{1-\rho_1^2}{1-2\rho_1 r_1 + r_1^2}\right)\left(\frac{1-r_1}{1-\rho_1}\right)^2$

ρ_1 \ r_1	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	1	0·80	0·62	0·45	0·31	—	—	—	—	—
0·1	1·22	1	0·78	0·58	0·41	0·27	—	—	—	—
0·2	1·50	1·25	1	0·76	0·54	0·36	0·21	—	—	—
0·3	1·86	1·58	1·29	1	0·73	0·49	0·30	0·16	—	—
0·4	2·33	2·03	1·70	1·35	1	0·69	0·42	0·23	0·09	—
0·5	—	2·67	2·29	1·86	1·42	1	0·63	0·34	0·14	0·03
0·6	—	—	3·20	2·68	2·12	1·54	1	0·55	0·24	0·05
0·7	—	—	—	4·14	3·40	2·58	1·74	1	0·44	0·10
0·8	—	—	—	—	6·23	5·00	3·60	2·19	1	0·24
0·9	—	—	—	—	—	13·57	10·86	7·43	3·80	1

Table 3. $f_l = \left(\frac{1-\rho_1^2}{1-2\rho_1 r_1 + r_1^2}\right)\left(\frac{1+r_1}{1+\rho_1}\right)^2$

ρ_1 \ r_1	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	1	1·20	1·38	1·55	1·69	—	—	—	—	—
0·1	0·82	1	1·18	1·34	1·48	1·60	—	—	—	—
0·2	0·67	0·83	1	1·16	1·31	1·43	1·52	—	—	—
0·3	0·54	0·69	0·84	1	1·15	1·28	1·38	1·45	—	—
0·4	0·43	0·56	0·70	0·85	1	1·13	1·25	1·33	1·39	—
0·5	—	0·44	0·57	0·71	0·86	1	1·12	1·22	1·29	1·32
0·6	—	—	0·45	0·58	0·72	0·87	1	1·11	1·19	1·24
0·7	—	—	—	0·45	0·58	0·72	0·87	1	1·10	1·16
0·8	—	—	—	—	0·42	0·56	0·71	0·87	1	1·08
0·9	—	—	—	—	—	0·34	0·48	0·66	0·85	1

Table 4. *Lower bound to efficiency of estimates.*

Case $p = 2, q = 0, p' = 1, q' = 0$; $\left|\frac{\lambda_1(1+\lambda_2)}{4\lambda_2}\right| \leqslant 1$

Parameters of the true error process		$\left(\frac{1-\rho_1^2}{1+\rho_1^2}\right)^2$	True lower bound to the efficiency
λ_1	λ_2		
−1·8	0·9	0·0030	0·0003
−0·1	0·9	0·9888	0·0099
−1·0	0·5	0·1476	0·0769
−0·5	−0·4	0·0327	0·0175
−0·1	−0·8	0·3600	0·0122

Table 5. *Lower bound to efficiency of estimates. Case* $p = 0, q = 2, p' = 0, q' = 1$. *Bracketed figures are* ρ_1 *and* ρ_2

μ_2 \ μ_1	0·2	0·5	0·8
−0·5	0·17 (0·08, −0·39)	0·00 (0·17, −0·33)	0·19 (0·21, −0·27)
−0·2	0·76 (0·15, −0·19)	0·43 (0·31, −0·16)	0·00 (0·38, −0·12)
0·2	0·79 (0·22, 0·19)	0·38 (0·47, 0·16)	—
0·5	0·28 (0·23, 0·39)	—	—
1·0	0·00 (0·20, 0·49)	0·00 (0·44, 0·44)	—
2·0	0·33 (0·12, 0·40)	0·26 (0·29, 0·38)	0·14 (0·43, 0·36)

Table 6. *Lower bound to efficiency of estimates. Case* $p = 0, q = 1, p' = 1, q' = 0$

r_1	0·2	0·2	0·2	0·2	0·4	0·4	0·4	0·4
μ_1	0·2	0·4	0·6	0·8	0·2	0·4	0·6	0·8
ρ_1	0·19	0·34	0·44	0·49	0·19	0·34	0·44	0·49
Minimum	0·99	0·81	0·42	0·10	0·81	0·90	0·55	0·15

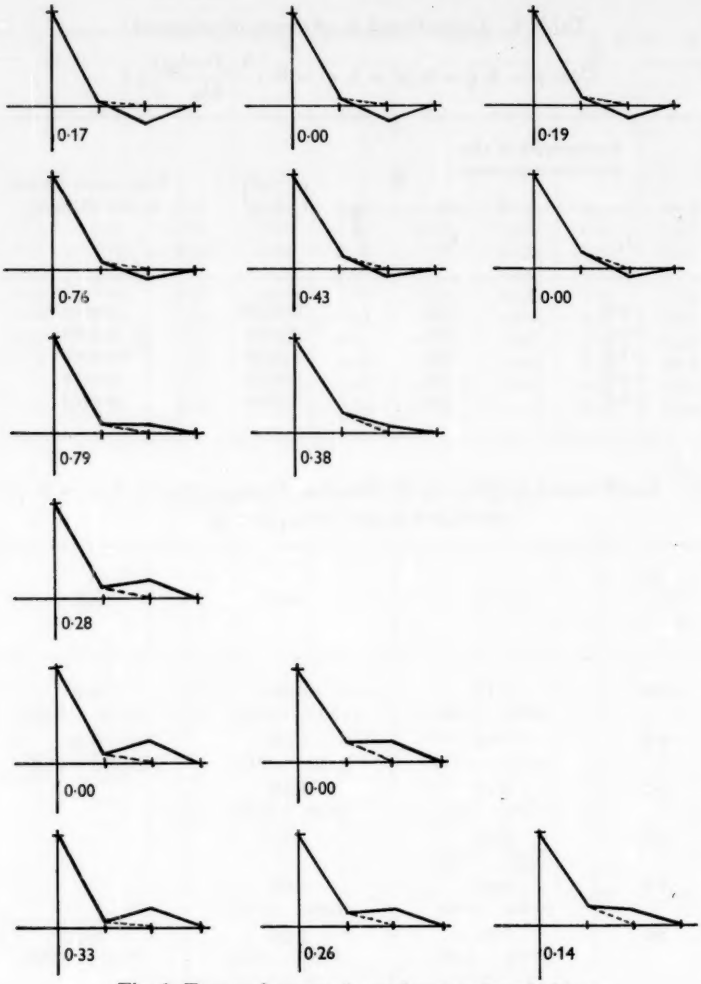


Fig. 1. True and assumed correlograms from Table 5.

Table 7. *Bounds to the* 5 % *significance level in t-tests*

Case	Exact (%)	First approximation (%)	Second approximation (%)
$N = 11$	1·4	1	2
$\rho_1 = 0{\cdot}3$, $r_1 = 0$	14·6	15	16
$N = 20$	—	2	4
$\rho_1 = 0{\cdot}7$, $r_1 = 0{\cdot}5$	21	20	20
$N = 20$	—	0·4	3
$\rho_1 = 0{\cdot}9$, $r_1 = 0{\cdot}5$	—	20	20

MISCELLANEA

Revised upper percentage points of the extreme studentized deviate from the sample mean

By H. A. DAVID

University of Melbourne

Let x_n and x_1 denote the largest and the smallest value respectively in a random sample of n observations drawn from a normal population with standard deviation σ. The extreme studentized deviate from the sample mean is defined as $(x_n-\bar{x})/s_\nu$ or $(\bar{x}-x_1)/s_\nu$, where s_ν is the usual root-mean-square estimator of σ based on ν degrees of freedom and is independent of the numerator. Tables of upper and lower 5 and 1% points and later also of 10, 2·5, 0·5 and 0·1% points were constructed by Nair (1948, 1952) for $n = 3$ to 9 and selected values of $\nu \geqslant 10$. The former set of tables is reproduced as Table 26 in Pearson & Hartley's *Biometrika Tables for Statisticians* (1954).

Below we give a revised and slightly extended table of the *upper* percentage points of $(x_n-\bar{x})/s_\nu$, which should be accurate except for possible errors of one unit in the last figure shown. Comparison with Nair's values makes it clear that the studentization procedure employed by him tends to become unsatisfactory for $\nu < 20$, as has been suspected by Pearson & Hartley (1954, p. 50, footnote). This effect becomes the more pronounced as n increases and the significance level α sharpens. For any given α the most severe correction is generally needed in the case $n = 9$, $\nu = 10$; in place of Nair's values 2·54, 2·93, 3·28, 3·67, 3·92, 4·40 for the upper 10, 5, 2·5, 1, 0·5, 0·1% points we find 2·50, 2·89, 3·29, 3·82, 4·24, 5·3.

The present table is based primarily on a simple approximation of a type recently discussed by the writer (David, 1956). Let $u_n = x_n-\bar{x}$. Since u_n is the largest of n deviations, $x-\bar{x}$, we have the approximate relation, valid for large values of R (constant),

$$\Pr(u_n/s_\nu > R) \doteqdot n \Pr[(x-\bar{x})/s_\nu > R].$$

It follows that the upper 100α% point of u_n/s_ν is given approximately by

$$[(n-1)/n]^{\frac{1}{2}}\, t_\nu(\alpha/n), \tag{1}$$

where $t_\nu(\alpha/n)$ is the upper $100\alpha/n$ percentage point of a t variate with ν degrees of freedom. As (1) slightly overestimates the true percentage points the approximate values were corrected by reference to a framework of exact percentage points constructed by numerical integration. These exact values were obtained as the solutions, R_α, of the equation

$$\int_0^\infty f(u_n) \int_0^{u_n/R_\alpha} f(s_\nu)\, ds_\nu\, du_n = \alpha.$$

Here $f(u_n)$, the probability density function of u_n, was found by differencing of Grubbs's (1950) table of the cumulative distribution function of u_n, and the inner integral evaluated from tables of the incomplete Γ-function. In the least favourable case ($\alpha = 0{\cdot}10$, $n = 12$, $\nu = 10$) the use of (1) led to an error of only 0·07.

I am much indebted to Misses Betty Laby and Gwenda Jacobs for extensive assistance with the computations.

REFERENCES

David, H. A. (1956). *Biometrika*, **43**, 85.
Grubbs, F. E. (1950). *Ann. Math. Statist.* **21**, 27.
Nair, K. R. (1948). *Biometrika*, **35**, 118.
Nair, K. R. (1952). *Biometrika*, **39**, 189.
Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

Table 1. *Upper percentage points of the extreme studentized deviate from the sample mean* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

ν \ n	3	4	5	6	7	8	9	10	12
	10 % points								
10	1·68	1·92	2·09	2·23	2·33	2·42	2·50	2·56	2·68
11	1·66	1·90	2·07	2·20	2·30	2·39	2·46	2·53	2·64
12	1·65	1·88	2·05	2·17	2·28	2·36	2·44	2·50	2·61
13	1·63	1·86	2·03	2·16	2·26	2·34	2·41	2·47	2·58
14	1·62	1·85	2·01	2·14	2·24	2·32	2·39	2·45	2·56
15	1·61	1·84	2·00	2·12	2·22	2·31	2·38	2·44	2·54
16	1·61	1·83	1·99	2·11	2·21	2·29	2·36	2·42	2·52
17	1·60	1·82	1·98	2·10	2·20	2·28	2·35	2·41	2·51
18	1·59	1·82	1·97	2·09	2·19	2·27	2·34	2·39	2·49
19	1·59	1·81	1·96	2·08	2·18	2·26	2·33	2·38	2·48
20	1·58	1·80	1·96	2·08	2·17	2·25	2·32	2·37	2·47
24	1·57	1·78	1·94	2·05	2·15	2·22	2·29	2·34	2·44
30	1·55	1·77	1·92	2·03	2·12	2·20	2·26	2·32	2·41
40	1·54	1·75	1·90	2·01	2·10	2·17	2·23	2·29	2·38
60	1·52	1·73	1·87	1·98	2·07	2·14	2·20	2·26	2·35
120	1·51	1·71	1·85	1·96	2·05	2·12	2·18	2·23	2·32
∞	1·50	1·70	1·83	1·94	2·02	2·09	2·15	2·20	2·28
	5 % points								
10	2·01	2·27	2·46	2·60	2·72	2·81	2·89	2·96	3·08
11	1·98	2·24	2·42	2·56	2·67	2·76	2·84	2·91	3·03
12	1·96	2·21	2·39	2·52	2·63	2·72	2·80	2·87	2·98
13	1·94	2·19	2·36	2·50	2·60	2·69	2·76	2·83	2·94
14	1·93	2·17	2·34	2·47	2·57	2·66	2·74	2·80	2·91
15	1·91	2·15	2·32	2·45	2·55	2·64	2·71	2·77	2·88
16	1·90	2·14	2·31	2·43	2·53	2·62	2·69	2·75	2·86
17	1·89	2·13	2·29	2·42	2·52	2·60	2·67	2·73	2·84
18	1·88	2·11	2·28	2·40	2·50	2·58	2·65	2·71	2·82
19	1·87	2·11	2·27	2·39	2·49	2·57	2·64	2·70	2·80
20	1·87	2·10	2·26	2·38	2·47	2·56	2·63	2·68	2·78
24	1·84	2·07	2·23	2·34	2·44	2·52	2·58	2·64	2·74
30	1·82	2·04	2·20	2·31	2·40	2·48	2·54	2·60	2·69
40	1·80	2·02	2·17	2·28	2·37	2·44	2·50	2·56	2·65
60	1·78	1·99	2·14	2·25	2·33	2·41	2·47	2·52	2·61
120	1·76	1·96	2·11	2·22	2·30	2·37	2·43	2·48	2·57
∞	1·74	1·94	2·08	2·18	2·27	2·33	2·39	2·44	2·52
	2·5 % points								
10	2·34	2·63	2·83	2·98	3·10	3·20	3·29	3·36	3·49
11	2·30	2·58	2·77	2·92	3·03	3·13	3·22	3·29	3·41
12	2·27	2·54	2·73	2·87	2·98	3·08	3·16	3·23	3·35
13	2·24	2·51	2·69	2·83	2·94	3·03	3·11	3·18	3·29
14	2·22	2·48	2·66	2·79	2·90	2·99	3·07	3·14	3·25
15	2·20	2·45	2·63	2·76	2·87	2·96	3·04	3·11	3·21
16	2·18	2·43	2·61	2·74	2·84	2·93	3·01	3·08	3·18
17	2·17	2·42	2·59	2·72	2·82	2·91	2·98	3·05	3·15
18	2·15	2·40	2·57	2·70	2·80	2·89	2·96	3·02	3·12
19	2·14	2·39	2·56	2·68	2·78	2·87	2·94	3·00	3·10
20	2·13	2·37	2·54	2·67	2·77	2·85	2·92	2·98	3·08
24	2·10	2·34	2·50	2·62	2·72	2·80	2·87	2·93	3·02
30	2·07	2·30	2·46	2·58	2·67	2·75	2·81	2·87	2·96
40	2·04	2·27	2·42	2·53	2·62	2·70	2·76	2·82	2·91
60	2·01	2·23	2·38	2·49	2·58	2·65	2·71	2·76	2·85
120	1·98	2·20	2·34	2·45	2·53	2·60	2·66	2·71	2·79
∞	1·95	2·16	2·30	2·41	2·49	2·56	2·61	2·66	2·74

Table 1 (*cont.*)

n \ ν	3	4	5	6	7	8	9	10	12
	1 % points								
10	2·78	3·10	3·32	3·48	3·62	3·73	3·82	3·90	4·04
11	2·72	3·02	3·24	3·39	3·52	3·63	3·72	3·79	3·93
12	2·67	2·96	3·17	3·32	3·45	3·55	3·64	3·71	3·84
13	2·63	2·92	3·12	3·27	3·38	3·48	3·57	3·64	3·76
14	2·60	2·88	3·07	3·22	3·33	3·43	3·51	3·58	3·70
15	2·57	2·84	3·03	3·17	3·29	3·38	3·46	3·53	3·65
16	2·54	2·81	3·00	3·14	3·25	3·34	3·42	3·49	3·60
17	2·52	2·79	2·97	3·11	3·22	3·31	3·38	3·45	3·56
18	2·50	2·77	2·95	3·08	3·19	3·28	3·35	3·42	3·53
19	2·49	2·75	2·93	3·06	3·16	3·25	3·33	3·39	3·50
20	2·47	2·73	2·91	3·04	3·14	3·23	3·30	3·37	3·47
24	2·42	2·68	2·84	2·97	3·07	3·16	3·23	3·29	3·38
30	2·38	2·62	2·79	2·91	3·01	3·08	3·15	3·21	3·30
40	2·34	2·57	2·73	2·85	2·94	3·02	3·08	3·13	3·22
60	2·29	2·52	2·68	2·79	2·88	2·95	3·01	3·06	3·15
120	2·25	2·48	2·62	2·73	2·82	2·89	2·95	3·00	3·08
∞	2·22	2·43	2·57	2·68	2·76	2·83	2·88	2·93	3·01
	0·5 % points								
10	3·12	3·46	3·70	3·87	4·02	4·14	4·24	4·33	4·47
11	3·04	3·37	3·59	3·76	3·90	4·01	4·11	4·19	4·33
12	2·98	3·29	3·51	3·67	3·80	3·91	4·00	4·08	4·21
13	2·93	3·23	3·44	3·60	3·72	3·83	3·92	3·99	4·12
14	2·88	3·18	3·38	3·54	3·66	3·76	3·85	3·92	4·04
15	2·84	3·13	3·33	3·48	3·60	3·70	3·78	3·86	3·98
16	2·81	3·10	3·29	3·44	3·56	3·65	3·73	3·80	3·92
17	2·78	3·07	3·26	3·40	3·52	3·61	3·68	3·75	3·86
18	2·76	3·04	3·23	3·37	3·48	3·57	3·64	3·71	3·82
19	2·74	3·01	3·20	3·34	3·45	3·54	3·61	3·68	3·79
20	2·72	2·99	3·17	3·31	3·42	3·51	3·58	3·65	3·75
24	2·66	2·92	3·10	3·23	3·33	3·42	3·49	3·55	3·65
30	2·60	2·86	3·03	3·15	3·25	3·33	3·40	3·46	3·55
40	2·55	2·79	2·96	3·08	3·17	3·25	3·31	3·37	3·46
60	2·50	2·73	2·89	3·01	3·10	3·17	3·23	3·28	3·37
120	2·45	2·67	2·83	2·94	3·02	3·09	3·15	3·20	3·28
∞	2·40	2·62	2·76	2·87	2·95	3·02	3·07	3·12	3·20
	0·1 % points								
10	4·0	4·3	4·6	4·8	5·0	5·2	5·3	5·4	5·6
11	3·8	4·2	4·5	4·7	4·8	5·0	5·1	5·2	5·3
12	3·7	4·1	4·3	4·5	4·7	4·8	4·9	5·0	5·1
13	3·6	4·0	4·2	4·4	4·5	4·6	4·7	4·8	5·0
14	3·5	3·9	4·1	4·3	4·4	4·5	4·6	4·7	4·9
15	3·5	3·8	4·0	4·2	4·3	4·4	4·5	4·6	4·8
16	3·4	3·7	4·0	4·1	4·3	4·4	4·5	4·5	4·7
17	3·4	3·7	3·9	4·1	4·2	4·3	4·4	4·5	4·6
18	3·3	3·6	3·9	4·0	4·1	4·2	4·3	4·4	4·5
19	3·3	3·6	3·8	4·0	4·1	4·2	4·3	4·4	4·5
20	3·3	3·6	3·8	3·9	4·0	4·1	4·2	4·3	4·4
24	3·2	3·5	3·7	3·8	3·9	4·0	4·1	4·2	4·3
30	3·1	3·4	3·6	3·7	3·8	3·9	4·0	4·0	4·1
40	3·0	3·3	3·5	3·6	3·7	3·7	3·8	3·9	4·0
60	2·9	3·2	3·4	3·5	3·6	3·6	3·7	3·8	3·8
120	2·9	3·1	3·3	3·4	3·5	3·5	3·6	3·6	3·7
∞	2·8	3·0	3·2	3·3	3·4	3·4	3·5	3·5	3·6

Exact linear sequential tests for the mean of a normal distribution

By J. TAYLOR
Food Research Department, Unilever Limited

1. Introduction

Wald (1947, chapter 7) described sequential probability ratio tests for the mean of a normal distribution of known variance. He gave formulae for their operating characteristics and mean sample size and, recognizing that these were approximate, also set limits to these quantities. For tests in which the mean sample size is small, i.e. in which the alternative hypotheses differ greatly, the errors of the first and second kind are much less than the nominal values. The mean sample size is greater than the nominal value and, although it is well beneath that of a fixed-sample-size test with the same actual errors, it may be close to that of such a test with the same nominal errors. Baker (1950) used random normal deviates to investigate the test for one pair of alternative hypotheses. For these, the error of the first kind was approximately the mean of Wald's limits and the mean sample size was approximately Wald's upper limit.

Sequential tests whose actual errors are close to pre-assigned values are desirable, and no practicable direct method of obtaining them has been proposed. It is suggested that they be obtained by calculating the errors of given tests and using interpolation. One way would be to use Baker's approximation, but this would depend on results for only one pair of alternative hypotheses. Another method would be to use an improved approximation to the operating characteristic, due to Page (1954*a*, *b*), and this is outlined in § 3 of this paper. The present author has used a method based on random samples to obtain tests for which the errors of the first and second kind are both 0·05, and the results of his investigation are presented in the remaining sections.

2. Notation

We shall assume that x is normally distributed with mean θ and standard deviation 1, and that the alternative hypotheses are

$$H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta = \theta_1.$$

The errors of the first and second kind are to be α and β.

3. Wald's test and Page's improved approximation

Wald's test consists of taking observations x_i in sequence and plotting

$$y_m = \sum_{i=1}^{m} x_i$$

against m (see Fig. 1). The boundaries L_0 and L_1 are the lines

$$\left.\begin{aligned} y_m &= \frac{1}{\theta_1-\theta_0}\ln\frac{\beta}{1-\alpha}+m\frac{\theta_0+\theta_1}{2} \\ \text{and}\qquad y_m &= \frac{1}{\theta_1-\theta_0}\ln\frac{1-\beta}{\alpha}+m\frac{\theta_0+\theta_1}{2} \end{aligned}\right\} \qquad (1)$$

respectively (ln denotes a natural logarithm).

The approximations in Wald's formulae are due to the fact that many schemes end some distance beyond L_0 and L_1 instead of on those lines (i.e. they end on one of the full lines down from the A_i, or on one of the broken lines up from the B_i, and not at points A_i or B_i). One consequence is that if equations (1) are used to calculate boundaries with, say, $\alpha = \beta = 0{\cdot}05$, the scheme will have errors which are somewhat less than 0·05.

Page (1954*a*) has shown how the true operating characteristic can be calculated by solution of an integral equation, and (1954*b*) has discussed methods of obtaining the solution. His notation corresponds to ours if we take

$$\theta = \tfrac{1}{2}(\theta_1-\theta_0),$$

and use the probabilities (as given in his Table 2) for $Z = \frac{1}{2}h$. These are the values of $\alpha = \beta$ for a scheme with boundaries

$$\left.\begin{aligned} y_m &= -Z + m\frac{\theta_0+\theta_1}{2} \\ \text{and} \qquad y_m &= Z + m\frac{\theta_0+\theta_1}{2}. \end{aligned}\right\} \tag{2}$$

The table is calculated for a given value of Z and may then be used to find the value of θ for which $\alpha = \beta$ has a desired value. Thus interpolation in Page's table shows that, for $Z = 5$, we shall obtain $\alpha = \beta = 0{\cdot}05$ for $\theta = 0{\cdot}26$, approximately. Similar calculations for other values of Z would give a curve connecting θ and Z for $\alpha = \beta = 0{\cdot}05$. Tests with $\alpha \neq \beta$ could be investigated by taking $Z \neq \frac{1}{2}h$.

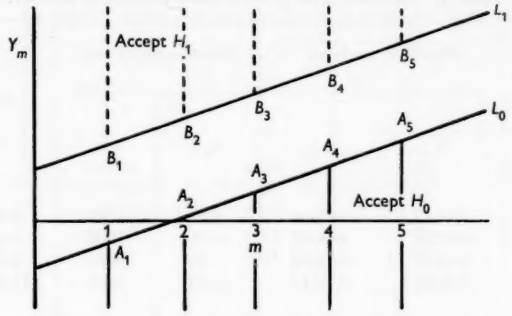


Fig. 1. Boundaries for a linear sequential scheme.

4. An alternative approach

Another approach is to fix $\theta_1 - \theta_0$ and calculate the values of $\alpha = \beta$ for various values of Z. There is no loss of generality if we take $\theta_0 = 0$ and the problem becomes that of determining the probability of crossing the upper boundary when the x_i are random normal deviates with mean 0 and standard deviation 1. Let P_{0i} (P_{1i}) be the probability that the lower (upper) boundary is crossed at $m = i$. We can calculate P_{01} and P_{11} from tables of the normal distribution; also P_{02} and P_{12} using the tetrachoric series (Kendall, 1941), as the joint distribution of y_1 and y_2 is bivariate normal with $\rho = \sqrt{0{\cdot}5}$. Further calculation is impracticable, but recourse may be made to the use of random sequences. The first 20,000 deviates in Wold's tables (1948) have been used to construct 2000 sequences of ten values ($y_1 \ldots y_{10}$). These have been used for several sets of (θ_1, Z) in which θ_1 is relatively large. Sequences which did not cross a boundary for $m \leqslant 10$ were completed from the remaining 5000 deviates in the tables. If N_c of the random sequences continued beyond $m = 2$ and U_c of these finally crossed the upper boundary, and if

$$P_c = 1 - P_{01} - P_{11} - P_{02} - P_{12},$$

then the estimated errors of the scheme are

$$\hat{\alpha} = \hat{\beta} = P_{11} + P_{12} + P_c\frac{U_c}{N_c},$$

and the standard error of $\hat{\alpha}$ is

$$P_c\sqrt{\left\{\frac{U_c}{N_c}\left(1 - \frac{U_c}{N_c}\right)\frac{1}{N_c}\right\}}.$$

Also if $\bar{n}_c$ is the mean sample size for these N_c sequences, with standard error s_c, the estimated mean sample size for the scheme as a whole (when H_0 or H_1 is true) is

$$n_{0,1} = (P_{01} + P_{11}) + 2(P_{02} + P_{12}) + P_c\bar{n}_c,$$

with standard error $P_c s_c$.

It may be noted that the proportion of sequences crossing the boundaries for $m = 1$ or 2 always agreed, within limits of sampling error, with the calculated values.

5. Results obtained by the sampling method

Table 1 gives the principal results. In order that efficiencies can be compared, the sizes n_f of fixed-sample-size tests with the same α, β and θ_1 are given as calculated, without rounding-off to the next highest integer.

For each value of θ_1 the probits of the values of $\hat{\alpha}$ are approximately linearly related to Z. Interpolation gives the values $\hat{Z}$ (in Table 2) for schemes having $\alpha = \beta = 0{\cdot}05$. The standard errors of the $\hat{Z}$ cannot be calculated, because the errors of the $\hat{\alpha}$ for different schemes are not independent (as the same random sequences were used for all schemes). However, we may be confident that the standard error of the significance level for each $\hat{Z}$ is under 0·005 and so the *true* significance level for each of the schemes given will be between 0·04 and 0·06. This degree of approximation is much less than in Wald's scheme.

Table 1. *Characteristics of some linear sequential schemes*

θ_1	Z	Significance level		Mean sample size		n_f	Efficiency $(n_f/n_{0.1})$
		$\hat{\alpha}$	S.E.	$n_{0.1}$	S.E.		
1·00	1·75	0·0988	0·0054	4·14	0·05	6·64	1·60
	2·00	0·0798	0·0052	4·71	0·06	7·91	1·68
	2·25	0·0647	0·0049	5·33	0·07	9·20	1·73
	2·50	0·0481	0·0044	6·03	0·08	11·07	1·84
1·25	1·50	0·0776	0·0043	3·15	0·03	5·17	1·64
	1·75	0·0593	0·0041	3·63	0·04	6·24	1·72
	2·00	0·0439	0·0038	4·15	0·05	7·44	1·79
1·50	1·00	0·0863	0·0031	2·06	0·02	3·31	1·61
	1·25	0·0643	0·0032	2·44	0·02	4·11	1·68
	1·50	0·0477	0·0032	2·83	0·03	4·95	1·75

Table 2. *Characteristics of interpolated schemes*

θ_1	$\hat{Z}$	Z_w	$\hat{Z}-Z_w$	$n_{0.1}$	n_f	Efficiency $(n_f/n_{0.1})$	$n_{\max.}$	S.E.
1·00	2·47	2·94	−0·47	5·93	10·82	1·82	9·18	0·32
1·25	1·90	2·36	−0·46	3·92	6·92	1·77	6·45	0·19
1·50	1·46	1·96	−0·50	2·76	4·81	1·74	4·43	0·12

Table 2 also gives values $$Z_w = \frac{\ln 19}{\theta_1},$$ calculated from Wald's equations (1) to correspond, for the given θ_1 and a nominal $\alpha = \beta = 0{\cdot}05$, to the Z in equations (2). The differences $\hat{Z} - Z_w$ show no regular trend and have a mean of $-0{\cdot}48$, so it is recommended that for other values of θ_1 in this range we should take

$$\hat{Z} = \frac{\ln 19}{\theta_1} - 0{\cdot}48.$$

The values $n_{0.1}$ have been interpolated from Table 1. They show that the efficiency of the schemes is good. The mean sample size when neither H_0 nor H_1 is true is also important, and will be a maximum ($n_{\max.}$) when $\theta = \frac{1}{2}\theta_1$. For each θ_1 considered, 500 random sequences were used to obtain the values of

$n_{\max.}$ and their standard errors. It will be seen that they are less than n_f. Thus these schemes may confidently be recommended, when they can be applied, as always having lower mean sample sizes than fixed-sample-size tests.

6. Truncation

It is a disadvantage of sequential schemes that they occasionally require very large sample sizes, and it is sometimes recommended that sampling should stop at two or three times the expected mean sample size. For the test given in this paper it would then be reasonable to accept H_0 or H_1 according to whether y_m was less than or greater than $\frac{1}{2}m\theta_1$ (i.e. above or below the mid-point of A_mB_m). This would give a test with

$$\alpha = \beta > \hat{\alpha}.$$

The mean sample size (n_t, say) would be less than $n_{0,1}$, but the increase in α and β reduces the size of the equivalent fixed-sample-size test. Table 3 shows the effect of truncating a particular linear scheme at various values of m.

Table 3. *Effect of truncation for* $\theta_1 = 1{\cdot}00$, $Z = 2{\cdot}50$

m	$\alpha\ (=\beta)$	n_t	n_f	Efficiency (n_f/n_t)
6	0·1244	4·50	5·32	1·18
9	0·0793	5·32	7·95	1·49
12	0·0654	5·70	9·13	1·60
15	0·0530	5·88	10·45	1·78
∞	0·0481	6·03	11·07	1·84

Thus truncation below $m = 15$ causes an appreciable increase in α and loss in efficiency. Similar calculations have been made for the other values of θ_1 and Z shown in Table 1, and it has been found generally true that truncation at $2{\cdot}5n_{0,1}$ is satisfactory, but not at lower values of m.

An alternative rule for action at the point of truncation would be to accept H_0 if the upper boundary had not been crossed, i.e. a path reaching any point on A_mB_m would result in the acceptance of H_0. This gives a guarantee that α is less than $\hat{\alpha}$, but β becomes greater than $\hat{\alpha}$ and the efficiency is appreciably less than for the symmetrical truncation considered above.

7. Summary

Wald's test for the mean of a normal distribution with known standard deviation involves an approximation which is poor when the alternative hypotheses differ greatly. The characteristics of some linear sequential tests have been calculated, and it is found that they have the high efficiency customary in sequential probability ratio tests. Some schemes for which the errors of the first and second kinds are 0·05 have been deduced. The truncation of such schemes has also been investigated and it has been concluded that this should not be done at less than 2·5 times the mean sample size on the null hypothesis.

Attention is drawn to a method, due to Page, of calculating the operating characteristic of Wald schemes. This provides another method of calculating schemes with prescribed errors.

The author thanks the directors of Unilever Limited for permission to publish this paper. He also thanks Miss R. M. Stimson, who did most of the computing involved.

REFERENCES

Baker, A. G. (1950). Properties of some tests in sequential analysis. *Biometrika*, **37**, 334–46.

Kendall, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalization. *Biometrika*, **32**, 196–8.

Page, E. S. (1954*a*). An improvement to Wald's approximation for some properties of sequential tests. *J. R. Statist. Soc.* B, **16**, 136–9.

Page, E. S. (1954*b*). The Monte Carlo solution of some integral equations. *Proc. Camb. Phil. Soc.* **50**, 414–25.

Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons, Inc.

Wold, H. (1948). *Tracts for Computers*. No. XXV. *Random Normal Deviates*. Cambridge University Press.

On the sum of squares of normal scores

BY H. RUBEN

Statistical Laboratory, Manchester University

Let $x_{r|n}$ $(r = 1, 2, \ldots, n)$ be the rth order statistic, in ascending order of magnitude, for a sample of n items drawn independently and at random from a standardized normal distribution. We shall obtain an explicit expression for $S_n = \sum_{r=1}^{n} a_{r|n}^2$, where $a_{r|n}$ is the expected value of $x_{r|n}$. The quantity S_n is of some importance in the analysis of ranked data and has been tabulated by Fisher & Yates (1953). It will appear that S_n can be expressed in terms of the contents of regular hyperspherical simplices with primary angles $\cos^{-1}(-\frac{1}{3})$ or equivalently in terms of the expectations of a set of *extreme* order statistics.

Define

$$a_{r|n} = n\binom{n-1}{r-1}\int_{-\infty}^{\infty} xp^{r-1}q^{n-r}z\,dx \quad (n = 1, 2, \ldots; r = 1, 2, \ldots, n), \tag{1}$$

where

$$z = \frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}x^2}, \tag{2}$$

$$p = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{1}{2}\xi^2}\,d\xi, \tag{3}$$

$$q = \frac{1}{\sqrt{(2\pi)}}\int_{x}^{\infty} e^{-\frac{1}{2}\xi^2}\,d\xi, \tag{4}$$

and

$$S_n = \sum_{r=1}^{n} a_{r|n}^2. \tag{5}$$

It is known (Ruben, 1954) that $a_{r|n}$ can be expressed in terms of the contents of hyperspherical simplices, constructed on the surfaces of spheres immersed in a space of dimensionality $n-2$, with primary angles either $\cos^{-1}(-\frac{1}{3})$ or $\pi - \cos^{-1}(-\frac{1}{3})$. In the particular case $r = 1$ or $r = n$ the simplices are regular, all primary angles then being $\cos^{-1}(-\frac{1}{3})$.

Let $u^{-(m)}$ and $u^{(m)}$ denote the ascending and descending factorials, respectively, of degree m in u, i.e.

$$\left.\begin{aligned} u^{-(0)} &= 1, \\ u^{-(m)} &= u(u+1)\ldots(u+m-1) \quad (m = 1, 2, \ldots), \end{aligned}\right\} \tag{6}$$

$$\left.\begin{aligned} u^{(0)} &= 1, \\ u^{(m)} &= u(u-1)\ldots(u-m+1) \quad (m = 1, 2, \ldots), \end{aligned}\right\} \tag{7}$$

and let $Q_{k,n}(r)$ denote the Tchebycheff–Hermite orthogonal polynomial of degree k in r for $r = 1, 2, \ldots, n$, which has been normalized to make the sum of squares of the values of each polynomial unity. Thus

$$Q_{k,n}(r) = K_{k,n}\sum_{j=0}^{k}(-)^{k-j}\frac{(k!)^2(k+j)!(n-j-1)!}{(2k)!(j!)^2(k-j)!(n-k-1)!}(r-1)^{(j)}$$
$$(k = 0, 1, \ldots, n-1;\ n = 1, 2, \ldots;\ r = 1, 2, \ldots, n) \tag{8}$$

with

$$K_{k,n} = \left[\frac{(2k)!(2k+1)!}{(k!)^4 n(n^2-1)\ldots(n^2-k^2)}\right]^{\frac{1}{2}}, \tag{8'}$$

and

$$\left.\begin{aligned} \sum_{r=1}^{n} Q_{k,n}(r)\,Q_{l,n}(r) &= 1 \quad (l=k) \\ &= 0 \quad (l\neq k) \end{aligned}\right\}. \tag{9}$$

From (9)

$$S_n = (Q_{0,n}(1)\,a_{1|n} + Q_{0,n}(2)\,a_{2|n} + \ldots + Q_{0,n}(n)\,a_{n|n})^2 + \ldots$$
$$+ (Q_{n-1,n}(1)\,a_{1|n} + Q_{n-1,n}(2)\,a_{2|n} + \ldots + Q_{n-1,n}(n)\,a_{n|n})^2. \tag{10}$$

Now $a_{r|n} = -a_{n-r+1|n}$ (this follows readily from (1)) and furthermore for *even* j, $Q_{j,n}(x)$ is even about $x = \frac{1}{2}(n+1)$. Equation (10) may therefore be written in the form

$$S_n = (Q_{1,n}(1)\,a_{1|n} + Q_{1,n}(2)\,a_{2|n} + \ldots + Q_{1,n}(n)\,a_{n|n})^2$$
$$+ (Q_{3,n}(1)\,a_{1|n} + Q_{3,n}(2)\,a_{2|n} + \ldots + Q_{3,n}(n)\,a_{n|n})^2 + \ldots$$
$$+ (Q_{m,n}(1)\,a_{1|n} + Q_{m,n}(2)\,a_{2|n} + \ldots + Q_{m,n}(n)\,a_{n|n})^2, \tag{10'}$$

where m is $n-1$ or $n-2$ according as to whether n is even or odd. To evaluate the separate terms on the right-hand side of equation (10′), note that

$$\begin{aligned}
\sum_{r=1}^{n} Q_{k,n}(r)\,a_{r|n} &= \sum_{j=0}^{k}\sum_{r=1}^{n} K_{k,n}\frac{(-)^{k-j}(k!)^2(k+j)!(n-j-1)!}{(2k)!(j!)^2(k-j)!(n-k-1)!}(r-1)^{(j)}n\binom{n-1}{r-1}\int_{-\infty}^{\infty} xp^{r-1}q^{n-r}z\,dx \\
&= nK_{k,n}\sum_{j=0}^{k}\frac{(-)^{k-j}(k!)^2(k+j)!(n-j-1)!}{(2k)!(j!)^2(k-j)!(n-k-1)!}\int_{-\infty}^{\infty}\left[\sum_{r=1}^{n}(r-1)^{(j)}\binom{n-1}{r-1}p^{r-1}q^{n-r}\right]xz\,dx \\
&= nK_{k,n}\sum_{j=0}^{k}\frac{(-)^{k-j}(k!)^2(k+j)!(n-j-1)!}{(2k)!(j!)^2(k-j)!(n-k-1)!}\int_{-\infty}^{\infty}(n-1)^{(j)}p^j xz\,dx \\
&= nK_{k,n}\int_{-\infty}^{\infty}\sum_{j=0}^{k}\frac{(-)^{k-j}(k+j)!(k!)^2}{(2k)!(k-j)!(j!)^2}(n-1)^{(k)}p^j xz\,dx \\
&= nK_{k,n}\sum_{j=0}^{k}\frac{(-)^{k-j}(k+j)!(k!)^2}{(2k)!(k-j)!(j!)^2}(n-1)^{(k)}j\frac{1}{\sqrt{(2\pi)}}u_{j-1}(\cos^{-1}-\tfrac{1}{3}),
\end{aligned} \tag{11}$$

where $u_{j-1}(\cos^{-1}-\frac{1}{3})$ is the relative content of a *regular* hyperspherical simplex constructed on the surface of a sphere immersed in $(j-1)$-dimensional space, the angle between any two bounding, diametral planes being $\cos^{-1}(-\frac{1}{3})$. The relationship between integrals of the form $\int_{-\infty}^{\infty} p^j xz\,dx$, and the u's has been demonstrated in a previous paper in which values of $u_m(\cos^{-1}1/x)$ have been tabulated for $x = 2(1)12$ and for various values of m (Ruben, 1954, Table 1). Also, in the derivation of (11) use has been made of the fact that the sum of the series in the square braces is the jth factorial moment of a random variable which has a binomial distribution with index $n-1$ and probability p; this factorial moment is $(n-1)^{(j)}p^j$.

On substitution of $K_{k,n}$ in (11) and squaring,

$$\left(\sum_{r=1}^{n} Q_{k,n}(r)\,a_{r|n}\right)^2 = n\frac{(n-1)^k}{(n-1)^{-(k)}}\alpha_k, \tag{12}$$

where

$$\alpha_k = \frac{2k+1}{4\pi}\left(\sum_{j=1}^{k}(-)^{k-j}\frac{(k+j)^{(2j)}}{j!(j-1)!}u_{j-1}(\cos^{-1}-\tfrac{1}{3})\right)^2 \quad (k=0,1,\ldots), \tag{13}$$

the even α's being all zero, as implied in the remarks made immediately after equation (10).

Hence, finally, S_n is expressed as a linear function of the first $[\frac{1}{2}n]$ odd α's:

$$\frac{1}{n}(a_{1|n}^2+\ldots+a_{n|n}^2) = \frac{n-1}{n+1}\alpha_1+\frac{(n-1)(n-2)(n-3)}{(n+1)(n+2)(n+3)}\alpha_3+\ldots. \tag{14}$$

Evidently since

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}(a_{1|n}^2+\ldots+a_{n|n}^2) &= E(x^2) \\
&= 1,
\end{aligned} \tag{15}$$

x being a standardized normal variate, we must have

$$\alpha_1+\alpha_3+\alpha_5+\ldots = 1. \tag{16}$$

We note that only the first two of the odd α's are expressible in terms of elementary functions. In fact, since

$$u_0(\theta) = 1, \quad u_1(\theta) = \tfrac{1}{2}, \quad u_2(\theta) = \theta/2\pi,$$

$$\alpha_1 = \frac{3}{\pi},$$

$$\alpha_3 = \frac{63}{\pi}\left(10\frac{\cos^{-1}(-\frac{1}{3})}{2\pi}-3\right)^2.$$

The reason for this is that α_3 involves the relative content of an arc of a circle (subtending an angle $\cos^{-1}(-\frac{1}{3})$ at the centre); on the other hand, $\alpha_5, \alpha_7, \ldots$ involve the relative contents of regular spherical tetrahedra, pentahedra, etc., and these cannot be expressed in terms of elementary trigonometric functions (cf. Ruben, 1954). The values of the first six odd α's are given below:

k	α_k
1	0·954929658728
3	0·033492010690
5	0·00667477593
7	0·002278473
9	0·0010182
11	0·00058

Fisher & Yates (1953, Table XXI) present S_n to 4 decimal places. These values were computed from the values of $a_{r|n}$, given to 2 decimal places in Table XX, by squaring and adding. An upper bound to the absolute error of the tabulated S_n is therefore* $2 \times 0{\cdot}005 \times \sum_{r=1}^{n} |a_{r|n}| \sim 0{\cdot}008n$ (about), since

$$\lim_{n\to\infty} \frac{1}{n}(|a_{1|n}| + \ldots + |a_{n|n}|) = E(|x|) = \sqrt{\frac{2}{\pi}},$$

x being again a standardized normal variate. Thus S_{50}, given as 47·3830 in the tables, may be in error by as much as 0·4, i.e. by somewhat less than 1 %.

The values of the α's tabulated above allow the computing of S_n for values of n from 1 to 13 inclusive with great precision by means of equation (14), the last term in the right-hand member of (14) involving α_{11} for $n = 12$ and $n = 13$. Upper bounds for the errors, induced by the errors of the tabulated α's, in S_{12} and S_{13} are $6{\cdot}1 \times 10^{-5}$ and $6{\cdot}6 \times 10^{-5}$ respectively, and are very much less for lower values of n. Thus the upper bounds for the errors in S_{10} and S_{11} are $5{\cdot}0 \times 10^{-7}$ and $5{\cdot}5 \times 10^{-7}$ respectively.

For values of $n > 13$ higher order α's are required if *all* the terms in the series of (14) are to be used. However, this is not strictly necessary, since high accuracy may be obtained by including terms as far as that involving α_{11} only. Fisher & Yates's tables may thus be extended to cover values of $n > 50$. The error in S_n now consists of a part due to the non-inclusion of terms involving $\alpha_{13}, \alpha_{15}, \ldots$ as well as (just as for $n \leqslant 13$) of a relatively small part due to the errors of the tabulated values of $\alpha_1, \ldots, \alpha_{11}$. An upper bound to the total error for S_{50} is 0·05, while for S_{100} it is 0·10. More generally for moderate or high values of n it is about 0·1 % of the value of S_n.

REFERENCES

FISHER, R. A. & YATES, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*. 4th ed. Edinburgh: Oliver and Boyd.

RUBEN, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200–27.

On the moments of the range and product moments of extreme order statistics in normal samples

BY H. RUBEN

Statistical Laboratory, Manchester University

It has been shown in a recent paper (Ruben, 1954) that the moments of order statistics in samples from normal populations are expressible as linear functions of the contents of certain hyperspherical simplices. This note will demonstrate that the product moments of the extreme order statistics in samples with an even number of items from normal populations are expressible as linear functions of the *products* of the contents of these simplices (equation (4)). The odd moments of the range when the number of items is odd, and the even moments when the number of items is even, may be expressed in a similar manner.

The joint probability density function of the extreme order statistics in normal samples of size n is

$$n(n-1)[F(v) - F(u)]^{n-2} f(u) f(v) \quad (v \geqslant u), \tag{1}$$

where

$$f(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}, \tag{2}$$

$$F(u) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{u} e^{-\frac{1}{2}\xi^2} d\xi. \tag{3}$$

* Actually, it is very slightly less for odd n, since $a_{r|n}$ is *accurately* zero when $r = \frac{1}{2}(n+1)$ and n is odd.

Hence, for n *even*,

$$\begin{aligned}
E(uv) &= n(n-1)\iint_{v\geqslant u}[F(v)-F(u)]^{n-2}f(u)f(v)\,uv\,du\,dv \\
&= \tfrac{1}{2}n(n-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[F(v)-F(u)]^{n-2}f(u)f(v)\,uv\,du\,dv \\
&= \tfrac{1}{2}n(n-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{n-2}(-)^k\binom{n-2}{k}[F(u)]^k[F(v)]^{n-2-k}f(u)f(v)\,uv\,du\,dv \\
&= \tfrac{1}{2}n(n-1)\sum_{k=0}^{n-2}(-)^k\binom{n-2}{k}\int_{-\infty}^{\infty}u[F(u)]^kf(u)\,du\int_{-\infty}^{\infty}v[F(v)]^{n-2-k}f(v)\,dv \\
&= \frac{1}{8\pi}n(n-1)\sum_{k=0}^{n-2}(-)^k\binom{n-2}{k}k(n-k-2)\,u_{k-1}(\cos^{-1}-\tfrac{1}{3})\,u_{n-k-3}(\cos^{-1}-\tfrac{1}{3})\quad(n=2,4,6,\ldots),
\end{aligned} \tag{4}$$

where $u_m(\theta)$ is the ratio of the content of a regular (hyper) spherical simplex, with primary angles θ, to the content of the surface of a sphere immersed in m-dimensional space, the simplex being constructed on the surface of such a sphere. In the derivation of equation (4), use has been made of the result (Ruben, 1954) that the expectation of the upper extreme statistic in a normal sample of size t is

$$t\int_{-\infty}^{\infty}u[F(u)]^{t-1}f(u)\,du=\frac{t(t-1)}{2\sqrt{\pi}}u_{t-2}(\cos^{-1}-\tfrac{1}{3})\quad(t=1,2,\ldots). \tag{5}$$

The relationship between the present u's and the $\overline{u}$'s of the previous paper is $u_m(\cos^{-1}-1/x)=\overline{u}_m(x)$. Equation (4) may be written in slightly simpler form, as follows:

$$\begin{aligned}
E(uv) &= \frac{n(n-1)}{8\pi}\sum_{k=1}^{n-3}(-)^k\binom{n-2}{k}k(n-2-k)\,u_{k-1}(\cos^{-1}-\tfrac{1}{3})\,u_{n-k-3}(\cos^{-1}-\tfrac{1}{3}) \\
&= \frac{n(n-1)}{8\pi}\sum_{k=1}^{n-3}(-)^k\frac{(n-2)!}{(k-1)!\,(n-k-3)!}u_{k-1}(\cos^{-1}-\tfrac{1}{3})\,u_{n-k-3}(\cos^{-1}-\tfrac{1}{3}) \\
&= -\frac{n(n-1)(n-2)(n-3)}{8\pi}\sum_{k=0}^{n-4}(-)^k\binom{n-4}{k}u_k(\cos^{-1}-\tfrac{1}{3})\,u_{n-k-4}(\cos^{-1}-\tfrac{1}{3})\quad(n=2,4,\ldots).
\end{aligned} \tag{6}$$

The terms in the series of equation (6) corresponding to k and to $n-4-k$ are equal so that yet another version of equation (4) more adapted to computation is

$$\begin{aligned}
E(uv) = &-\frac{n(n-1)(n-2)(n-3)}{4\pi}\sum_{k=0}^{\frac{1}{2}n-3}(-)^k\binom{n-4}{k}u_k(\cos^{-1}-\tfrac{1}{3})\,u_{n-4-k}(\cos^{-1}-\tfrac{1}{3}) \\
&-\frac{n(n-1)}{8\pi}(-)^{\frac{1}{2}n}\binom{n-2}{\frac{1}{2}(n-2)}(\tfrac{1}{2}n-1)^2u^2_{\frac{1}{2}n-2}(\cos^{-1}-\tfrac{1}{3})\quad(n=2,4,\ldots).
\end{aligned} \tag{7}$$

A similar argument may be used to derive $E(v-u)^m$, the moments of the range w_n based on a sample of size n from normal populations, when m and n are both odd and when m and n are both even, since the integrands defining these moments will then be symmetrical in u and v as in the derivation of equation (4). The moments themselves may then be expressed as sums of products of contents of simplices whose angles are either θ or $\pi-\theta$ where

$$\theta=\cos^{-1}(-\tfrac{1}{3}),\quad\cos^{-1}(-\tfrac{1}{4}),\ldots,\cos^{-1}\left(-\frac{1}{m+2}\right).$$

We shall not derive the relevant expressions here but shall instead use (4) to obtain the second moment only of w_n for even n. We have

$$\begin{aligned}
E(w_n^2) &= E(v-u)^2 \\
&= 2E(v^2)-2E(uv) \\
&= 2+\frac{n(n-1)(n-2)}{2\pi\sqrt{3}}u_{n-3}(\cos^{-1}-\tfrac{1}{4}) \\
&\quad+\frac{n(n-1)(n-2)(n-3)}{4\pi}\sum_{k=0}^{n-4}(-)^k\binom{n-4}{k}u_k(\cos^{-1}-\tfrac{1}{3})\,u_{n-4-k}(\cos^{-1}-\tfrac{1}{3})\quad(n=2,4,\ldots),
\end{aligned} \tag{8}$$

on using equation (101′) of an earlier paper (Ruben, 1954) to express $E(v^2)$ in the form used in equation (8).

It should be noted incidentally that for *all* n

$$\begin{aligned} E(w_n) &= E(v-u) \\ &= 2E(v) \\ &= \frac{n(n-1)}{\sqrt{\pi}} u_{n-2}(\cos^{-1}-\tfrac{1}{3}) \quad (n = 1, 2, \ldots), \end{aligned} \tag{9}$$

on using equation (101′) referred to earlier.

Equations (8) and (9) may be used to devise explicit expressions for the variances of w_2, w_4 and w_6 in terms of elementary trigonometric functions. Similar expressions for the variance of w_n when $n = 8, 10, \ldots$ are not possible, since the formulae for these variances involve the contents of spherical tetrahedra, pentahedra, etc., and these cannot be expressed in terms of the elementary trigonometric functions (cf. Ruben, 1954, 1956). On the other hand, $E(w_6^2)$, as opposed to $\operatorname{var} w_6$, can be expressed in the desired form. In fact,

$$\begin{aligned} E(w_2) &= \frac{2}{\sqrt{\pi}} u_0(\cos^{-1}-\tfrac{1}{3}) \\ &= \frac{2}{\sqrt{\pi}}, \end{aligned} \tag{10}$$

$$\begin{aligned} E(w_4) &= \frac{12}{\sqrt{\pi}} u_2(\cos^{-1}-\tfrac{1}{3}) \\ &= \frac{12}{\sqrt{\pi}} \frac{\cos^{-1}-\frac{1}{3}}{2\pi}, \end{aligned} \tag{11}$$

$$E(w_6) = \frac{30}{\sqrt{\pi}} u_4(\cos^{-1}-\tfrac{1}{3}), \tag{12}$$

$$E(w_2^2) = 2, \tag{13}$$

$$\begin{aligned} E(w_4^2) &= 2 + \frac{24}{2\pi\sqrt{3}} u_1(\cos^{-1}-\tfrac{1}{4}) + \frac{24}{4\pi} u_0^2(\cos^{-1}-\tfrac{1}{3}) \\ &= 2 + \frac{6}{\pi\sqrt{3}} + \frac{6}{\pi}, \end{aligned} \tag{14}$$

$$\begin{aligned} E(w_6^2) &= 2 + \frac{120}{2\pi\sqrt{3}} u_3(\cos^{-1}-\tfrac{1}{4}) + \frac{360}{4\pi}\{2u_0(\cos^{-1}-\tfrac{1}{3})\, u_2(\cos^{-1}-\tfrac{1}{3}) - 2u_1^2(\cos^{-1}-\tfrac{1}{3})\} \\ &= 2 + \frac{120}{2\pi\sqrt{3}} \frac{3(\cos^{-1}-\frac{1}{4})-\pi}{4\pi} + \frac{360}{4\pi}\left\{2\cdot 1 \frac{\cos^{-1}-\frac{1}{3}}{2\pi} - 2(\tfrac{1}{2})^2\right\} \\ &= 2 - \frac{15}{\pi\sqrt{3}} - \frac{45}{\pi} + \frac{15}{\pi^2}\{\sqrt{3}\cos^{-1}(-\tfrac{1}{4}) + 6\cos^{-1}(-\tfrac{1}{3})\}, \end{aligned} \tag{15}$$

$$\operatorname{var} w_2 = 2 - \frac{\pi}{4}, \tag{16}$$

$$\operatorname{var} w_4 = 2 + \frac{6}{\pi} + \frac{2\sqrt{3}}{\pi} - \frac{36}{\pi^2}(\cos^{-1}(-\tfrac{1}{3}))^2. \tag{17}$$

Work is currently proceeding on the Manchester electronic computing machine for the evaluation of $\bar{u}_n(x)$ and $V_{r,n}(x)$ (the latter are related to the contents of skew spherical simplices, $\bar{u}_n(x) \equiv V_{n,n}(x)$ (Ruben, 1954)), at least as far as $n = 100$, $x = 6$ and for all possible r. At the same time the first four moments of all order statistics in normal samples of size $n \leqslant 100$ are being computed.

It is also hoped shortly to compute the product moments and hence the covariances of all order statistics for normal samples over the same range of n.

REFERENCES

RUBEN, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200.

RUBEN, H. (1956). On the sum of squares of normal scores. *Biometrika*, **43**, 456.

On estimating binomial response relations

By F. J. ANSCOMBE

*Statistical Laboratory, University of Cambridge**

Berkson, in numerous papers, has rightly stressed the preferability of assuming a logistic form of dose-response law to assuming an integrated-normal law: (i) the logistic law has, in some applications at least, some sort of theoretical justification, while the integrated-normal law has none, apart from a wholly speculative argument about a distribution of individual tolerances; (ii) the shapes of the two types of response curve are so similar that only a very extensive series of observations could hope to show any appreciable difference in goodness of fit; (iii) simple sufficient statistics are available for the parameters of the logistic law.

The objects of this note are (*a*) to point out that fitting a logistic law by maximum likelihood is not as laborious as Finney, in several publications, has made it out to be, and (*b*) to examine Berkson's method of fitting by 'minimum logit χ^2', suggesting a small modification of it. Remarks are also made about testing the adequacy of the assumed logistic form of law, and about fitting the 'angular' response law.

1. Notation

Groups of subjects are tested at each of k different values x_i of a predetermined variable ($i = 1, 2, \ldots, k$). In toxicology, x_i would be the ith dose expressed logarithmically; and it will be convenient here to refer to the x's as 'doses'. Of the n_i subjects tested at dose x_i, r_i are observed to respond. It is assumed that the r_i have independent binomial distributions with parameters n_i and P_i ($= 1-Q_i$), where P_i is a given function of unknown parameters α and β requiring estimation. The logistic law is that

$$\left.\begin{aligned} P_i &= [1+\exp(-\alpha-\beta x_i)]^{-1}, \\ \text{or} \qquad \ln(P_i/Q_i) &= \alpha+\beta x_i. \end{aligned}\right\} \tag{1}$$

It is also convenient to write $\alpha+\beta x_i = \beta(x_i-\mu)$, so that μ ($= -\alpha/\beta$) is the dose for which $P = \frac{1}{2}$.

Often there are two or more such sets of observations, and then it may be appropriate to assume that the parameter β is the same for each set. For simplicity of discussion, however, it will be supposed now that only one series of observations is being considered.

Since all symbols for quantities other than the parameters α, β, μ carry the suffix i, it will be convenient to omit the suffix, noting that all summations are over $i = 1, 2, \ldots, k$. Estimates of α and β will be denoted by $\hat{\alpha}$ and $\hat{\beta}$, and values of P and Q obtained from (1) when α and β are replaced by $\hat{\alpha}$ and $\hat{\beta}$ will be denoted by $\hat{P}$ and $\hat{Q}$.

2. Fitting the logistic law by maximum likelihood

Since Σr and Σrx are sufficient statistics and simple to calculate, it seems reasonable to use them. The maximum-likelihood equations for $\hat{\alpha}$ and $\hat{\beta}$ are (as Berkson has pointed out)

$$\Sigma r = \Sigma n\hat{P}, \quad \Sigma rx = \Sigma nx\hat{P}. \tag{2}$$

They can be solved iteratively as follows (cf. Cornfield, 1954). Given trial values of $\hat{\alpha}$ and $\hat{\beta}$, calculate the right-hand sides of (2) and also the estimated weights $\hat{W}$ defined by

$$\hat{W} = n\hat{P}\hat{Q}. \tag{3}$$

Corrections $\delta\hat{\alpha}$ and $\delta\hat{\beta}$ to the trial values are then found by solving the equations

$$\left.\begin{aligned} (\Sigma\hat{W})\,\delta\hat{\alpha}+(\Sigma\hat{W}x)\,\delta\hat{\beta} &= \Sigma r-\Sigma n\hat{P}, \\ (\Sigma\hat{W}x)\,\delta\hat{\alpha}+(\Sigma\hat{W}x^2)\,\delta\hat{\beta} &= \Sigma rx-\Sigma nx\hat{P}. \end{aligned}\right\} \tag{4}$$

In further iterations it is unnecessary to recalculate the coefficients on the left-hand sides of (4), but only the right-hand sides. This procedure is to be compared with Finney's (e.g. 1952), which is exactly modelled on the ingenious procedure due to Fisher and Bliss for fitting the integrated-normal law, for which no such sufficient statistics are available.

Under the conditions that all the n's are equal, and that the x's are equally and not too widely spaced, and cover such a wide range that PQ may be assumed effectively to vanish outside it, the sums on the

* Present address: Department of Mathematics, Princeton University.

right-hand sides of (2) may be replaced by integrals, with the aid of the Euler–Maclaurin formula. After integration by parts we obtain

$$\left.\begin{aligned}\hat{\mu} &= x_k + \tfrac{1}{2}h - h\Sigma r/n,\\ \hat{\mu}^2 + \pi^2/(3\hat{\beta}^2) &= (x_k + \tfrac{1}{2}h)^2 - \tfrac{1}{12}h^2 - 2h\Sigma rx/n,\end{aligned}\right\} \tag{5}$$

where x_k is the greatest dose and h is the spacing between adjacent doses. The first of these equations is the Spearman–Kärber formula for estimating μ, the full efficiency of which was noted by Cornfield & Mantel (1950). The left-hand side of the second equation is approximately $\hat{\mu}^2 + 3{\cdot}290/\hat{\beta}^2$.

3. 'Minimum logit χ^2' (Berkson, 1953)

For graphical treatment it is natural to consider the transform of r by the inverse of the logistic function, namely,

$$l = \ln \frac{r}{n-r}. \tag{6}$$

This may be expected to show an approximately linear regression on x. Berkson has proposed a non-iterative method of fitting which consists of choosing $\hat{\alpha}$ and $\hat{\beta}$ to minimize

$$\Sigma W(l - \hat{\alpha} - \hat{\beta}x)^2, \tag{7}$$

where the W's are empirical weights defined by

$$W = r(n-r)/n. \tag{8}$$

(The procedure is modified a little, however, if there are any doses for which $r = 0$ or n.) It is easy to see that in large samples this is equivalent to the maximum-likelihood method of fitting. 'In large samples' means that, the x's being fixed, all the n's $\to \infty$.

Now as an estimate of $\alpha + \beta x$ the l defined by (6) is biased, but the bias can be very nearly removed by redefining l thus:

$$l = \ln \frac{r + \frac{1}{2}}{n - r + \frac{1}{2}}. \tag{9}$$

Moments of the asymptotic distribution of this l may be found by expanding the right-hand side of (9) as a Taylor series about the value nP for r, raising it to a power, and then taking expectations term by term. We find in this way

$$\mathscr{E}(l) = \alpha + \beta x + O(n^{-2}), \quad \operatorname{var}(l) \sim \frac{1}{nPQ}, \quad \gamma_1(l) \sim \frac{2(P-Q)}{\sqrt{(nPQ)}}. \tag{10}$$

The last of these results implies that the skewness of the distribution of l is asymptotically twice that of the binomial distribution of r and is of the opposite sign.

It seems, in fact, that we may accept the asymptotic mean and variance as good approximations, provided n is fairly large and P is not very close to 0 or 1. Since, for given n, l cannot lie outside $\pm \ln(2n+1)$, l is valueless as an estimate of $\alpha + \beta x$ if P is sufficiently close to 0 or 1, i.e. if $|x - \mu|$ is large enough. For any numerical values of n and P, it is possible to calculate $\mathscr{E}(l - \alpha - \beta x)$ and $\operatorname{var}(l)$ by summation over the binomial distribution of r. If n is not too small, it will be sufficient to consider the values for given nPQ, in the limit $n \to \infty$, and either P or $Q \to 0$. The formulae needed are then

$$\pm\, \mathscr{E}(l - \alpha - \beta x) = e^{-nPQ}\left\{nPQ \ln 3 + \frac{1}{2!}(nPQ)^2 \ln 5 + \frac{1}{3!}(nPQ)^3 \ln 7 + \ldots\right\} - \ln(2nPQ), \tag{11}$$

where the upper sign on the left-hand side refers to $P \to 0$, the lower one to $Q \to 0$, and

$$\operatorname{var}(l) = e^{-nPQ}\left\{nPQ(\ln 3)^2 + \frac{1}{2!}(nPQ)^2 (\ln 5)^2 + \ldots\right\} - e^{-2nPQ}\left\{nPQ \ln 3 + \frac{1}{2!}(nPQ)^2 \ln 5 + \ldots\right\}^2. \tag{12}$$

Some values obtained are shown in Table 1. The bias is positive if P is close to zero, negative if Q is close to zero. Bias and variance are to be compared with the variance $1/nPQ$ assumed in the least-squares fitting. It will be seen that when $nPQ = 2$ the bias is slight and the variance is close to $1/nPQ$; while if $nPQ = \frac{1}{2}$ or less, the mean-square error is much less than $1/nPQ$ but is largely due to the bias. If there are two or more doses for which $nPQ < 1$, the observations may well do more harm than good if they are given positive weights in the least-squares estimation of α and β.

It seems to me advantageous to try to eliminate the bias in l by using the definition (9) rather than (6), if the method of least squares is to be used, since that leads us to calculate (weighted) sums of the l's. Suppose there are only two doses, and the corresponding values of nPQ exceed 2. Then provided the

weights W that are assigned are positive, their values are immaterial. The fitted regression line passes through the points (x_1, l_1) and (x_2, l_2), and $\hat{\alpha}$ and $\hat{\beta}$ are unbiased. If now there are three such doses, roughly equally spaced, the fitted regression line will depend on the weights, but not much; it will make little difference whether empirical weights (8) are used, or fitted weights (3).

The precise definition of the weight function (so long as it approximates to nPQ) only begins to matter if (*a*) there are several doses for which either P or Q is very small, or (*b*) there are a considerable number of closely spaced doses for which neither P nor Q is very small. If (*a*), the bias in l when nPQ is small will matter if $W \neq 0$; if (*b*), a purely empirical weight such as (8) will introduce a bias, the extent of which will depend on how close the doses are. It seems likely that no purely empirical weight function such as (8) can be satisfactory for all possible sets of doses, though it may well be that the procedure explained in detail by Berkson (1953) (and *not* fully explained here) is satisfactory for all ordinary use where the doses are roughly equally spaced and neither too close nor too far apart.

To guard against the effect of unusual spacing of doses, I have formulated the following recommendation (or rather, two alternative recommendations), which should yield estimates of α and β that are almost unbiased and of minimum variance.

Table 1. *Moments of the distribution of l defined by* (9)

nPQ	Magnitude of bias	Variance	Mean-square error
2	0·026	0·473	0·474
$\sqrt{2}$	0·078	0·515	0·521
1	0·169	0·505	0·534
$\frac{1}{2}$	0·484	0·385	0·619
$\frac{1}{4}$	0·950	0·240	1·144

Suggested modification of Berkson's method. *Defining l by* (9), *minimize the sum of squares* (7), *where* either W *is defined by* (8) or W *is replaced by a fitted weight* $\hat{W}$ *equal to* $n\hat{P}\hat{Q}$ *if that exceeds* 1 *and equal to* 0 *otherwise.*

The empirical weight (8) would be used in cases where it is judged that the weights do not matter much anyway, as just explained. Otherwise, the fitted weight would be used; that involves the possibility of further iterations, but the process will converge quicker than the method of maximum likelihood because the iteration only concerns the weights, not the ordinates. The suggested critical value of 1 for $n\hat{P}\hat{Q}$ is, of course, rough-and-ready—any value between 2 and $\frac{1}{2}$ might be satisfactory. Unless there are at least two doses for which positive weights may confidently be assigned, the method must be considered to have broken down, and no estimates should be derived.

Whether the above procedure is preferable to the method of maximum likelihood is a question that cannot be answered without reference to the object in view. Granted the assumptions of independent binomial observations and a logistic response law, the fullest possible summary of the information contained in the observations is provided by the whole likelihood function. For some purposes, it suffices to quote merely the position of the maximum of the function, with the values of the second derivatives there. Berkson's procedure may be valuable as a substitute for this, or as a step towards it. Since, however, the assumptions cannot be asserted absolutely, a doubt concerning their appropriateness must always lurk; and graphical presentation of the data by plotting l against x has value as a check, apart from any approximation to the maximum-likelihood estimates.

4. Testing goodness of fit

When the question of goodness of fit of an assumed response law arises, it has been the practice to calculate a goodness-of-fit χ^2 (e.g. Armitage & Allen, 1950). This is rather unsatisfactory, since χ^2 will be sensitive to departures from the binomial distribution of responses at any dose, and often in practice one cannot be confident that the binomial distribution is very close to the truth. In any case, χ^2 does not indicate the direction of departure.

A better procedure, for testing the adequacy of the logistic law, would be to estimate parameters β_2 and β_3, assumed small, where

$$\ln(P/Q) = \beta_1(x-\mu) + \beta_2(x-\mu)^2 + \beta_3(x-\mu)^3. \qquad (13)$$

The appropriate statistics, besides Σr and Σrx already used in fitting the logistic law, are Σrx^2 and Σrx^3 (or similar expressions with $\hat{W}l$ instead of r). There is not much chance of detecting anything interesting unless there are a considerable number of series of observations (in the field of research being considered), and for each series the doses satisfy the conditions mentioned above for the Spearman–Kärber method to be valid—especially the condition of covering a wide range. It is then easy to estimate β_2 and β_3 for each series, by equations similar to (5).

If $\beta_2 \neq 0$, the response curve of P against x is not antisymmetrical about $x = \mu$. This situation might be remedied by a transformation of the x-scale. If $\beta_2 = 0$ and $\beta_3 \neq 0$, the logistic law is not satisfactory, but possibly one of the other suggested laws might be (see Finney, 1952, chap. 17). If estimates of β_2 and β_3 are available from a number of assays, their signs, rather than their values, will be of particular interest.

5. Fitting the angular law

The only other law besides the logistic which seems at present to be worth considering is

$$\left.\begin{aligned} P &= \sin^2(\alpha+\beta x) \quad \text{if} \quad 0<\alpha+\beta x<\tfrac{1}{2}\pi, \\ &= 0 \quad \text{otherwise.} \end{aligned}\right\} \tag{14}$$

Here the weight is constant (and positive) for $\alpha+\beta x$ in $(0, \frac{1}{2}\pi)$ and zero elsewhere. For this law, as for the logistic, the transform of r by the inverse response function can be modified so that it is asymptotically unbiased, whatever the value of $\alpha+\beta x$ in the range $(0, \frac{1}{2}\pi)$. The modified transformation, corresponding to (9) above, is (Anscombe, 1954)

$$y = \sin^{-1}\sqrt{\frac{r+\frac{1}{4}}{n+\frac{1}{2}}}. \tag{15}$$

Moments, corresponding to (10), are

$$\mathscr{E}(y) = \alpha+\beta x+O(n^{-2}), \quad \operatorname{var}(y) \sim \frac{1}{4n}, \quad \gamma_1(y) \sim \frac{P-Q}{2\sqrt{(nPQ)}}. \tag{16}$$

The device of adding equal constants to r and $n-r$ so that the transformed quantity is approximately unbiased, for all x in an appropriate range, is possible only for response laws of a certain mathematical type, such that x can be expressed as an integral with respect to P of a negative power of $P(1-P)$. The device is not therefore available for the integrated-normal response law.

I have taken advantage of helpful comments by Dr P. Armitage, Dr J. Berkson and Dr J. Cornfield in revising the draft of this paper.

Added in proof. In §3 above it is pointed out that (i) there seems to be no particular reason why the transformed quantity l should be defined by (6) rather than by some less obvious expression such as (9), (ii) the properties of l can be, and need to be, investigated, and (iii) if n is large the definition (9) is superior to (6), and seems satisfactory. Nothing has been said about the minimum logit χ^2 method when n is small; the results given might be expected to apply if $n=50$, but hardly if $n=5$. Even if n is large, (9) is not the best possible definition; Professor J. W. Tukey has pointed out that (9) can be improved by adding $\frac{1}{2}$ to the right-hand side of (9) when $r=n$, and subtracting $\frac{1}{2}$ when $r=0$, so that the range of values assumed by l is widened a little. No doubt by numerical study it would be possible to improve substantially on (9) when n is small.

REFERENCES

Anscombe, F. J. (1954). Comments on a paper by R. A. Fisher. *Biometrics*, **10**, 141–4.

Armitage, P. & Allen, I. (1950). Methods of estimating the LD 50 in quantal response data. *J. Hyg., Camb.*, **48**, 298–322.

Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. Statist. Ass.* **48**, 565–99.

Cornfield, J. (1954). Measurement and comparison of toxicities: the quantal response. *Statistics and Mathematics in Biology*, pp. 327–44. Ames: Iowa State College Press.

Cornfield, J. & Mantel, N. (1950). Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve. *J. Amer. Statist. Ass.* **45**, 181–210.

Finney, D. J. (1952). *Statistical Method in Biological Assay*. London: Griffin.

Existence and uniqueness of a uniformly most powerful randomized unbiased test for the binomial

By A. A. BLANK

Department of Mathematics, University of Tennessee

Introduction

Tocher (1950) applied the Neyman–Pearson theory of testing to discrete variables. In an example he pointed out the fact that among inbiased tests for the binomial those which are most powerful possess a certain special form. The questions of uniqueness and existence were left open.

Let θ denote the probability of success in a binomial trial. The problem is to test the hypothesis $\theta = p \in (0, 1)$ against the alternative $\theta \neq p$ by performing a sequence of n binomial trials. Let X be a random variable denoting the number of successes in the n trials. A test is equivalent to the choice of an acceptance criterion $0 \leqslant \psi(X) \leqslant 1$ defined so that the hypothesis $\theta = p$ is accepted with probability $\psi_\nu = \psi(\nu)$ whenever ν is the number of successes ($\nu = 0, 1, 2, \ldots, n$). The probability of accepting the hypothesis in the event of its truth is then

$$\sum_{\nu=0}^{n} \psi_\nu b_\nu(p) = 1-\alpha, \tag{1}$$

where $b_\nu(p)$ is the binomial term $\binom{n}{\nu} p^\nu (1-p)^{n-\nu}$ and α is the *size* of the test. Such a test is said to be *unbiased* if the probability of accepting the hypothesis when false is less than the probability of acceptance when true, that is,

$$\sum_{\nu=0}^{n} \psi_\nu b_\nu(\theta) \leqslant 1-\alpha \quad (0 \leqslant \theta \leqslant 1). \tag{2}$$

A test is said to be *uniformly most powerful* if simultaneously for all alternatives $\theta \neq p$ the probability of falsely accepting the hypothesis is minimized. Tocher showed that if a uniformly most powerful test exists among unbiased size α tests it must have the form

$$\left.\begin{aligned} &\psi_\nu = 1 \quad \text{for} \quad s+1 \leqslant \nu \leqslant t-1, \\ &\psi_s = c, \quad \psi_t = d, \\ &\psi_\nu = 0 \quad \text{otherwise.} \end{aligned}\right\} \tag{3}$$

It will be shown that among tests of this form there exists a unique unbiased test. Since Tocher's analysis may be applied directly to show this is most powerful among unbiased tests the proof will then be complete.

Proof of existence and uniqueness

Let us assume that a test (3) is given. We seek a value $\theta = p$ which minimizes the size α and maximizes the acceptance function

$$P(\theta) = \sum_{0}^{n} \psi_k b_k(\theta) = c b_s(\theta) + d b_t(\theta) + \sum_{s+1}^{t-1} b_k(\theta)$$

in the interval $0 \leqslant \theta \leqslant 1$, where we assume

$$0 < c,\ d < 1,$$
$$0 \leqslant s < t \leqslant n.$$

The effect of the first assumption is to guarantee that the terms of index s and t are present. The second is not restrictive since the proof is direct when $s = t$.

We have

$$\begin{aligned} P'(\theta) &= \sum_{s}^{t} \psi_k b_k(\theta) \frac{k-n\theta}{\theta(1-\theta)} \\ &= \frac{1}{\theta(1-\theta)} [S b_s - T b_t], \end{aligned}$$

where

$$\begin{aligned} S &= s(1-\theta) + (1-c)(n\theta - s), \\ T &= t(1-\theta) - d(t-n\theta) \\ &= t(1-d) + \theta(nd-t). \end{aligned}$$

The derivative $P'(\theta)$ can vanish only if (i) $\theta = 0$, (ii) $\theta = 1$, or (iii) θ lies in the open interval $I = (s/n, t/n)$.

Consider first the case $s = 0$, $t = n$. If $c = d = 1$, $P(\theta) \equiv 1$ and the case is trivial. Otherwise p must satisfy $\{p/(1-p)\}^{n-1} = (1-c)/d$, which has a unique solution since $\theta/(1-\theta)$ is monotone and unbounded. A unique extremum exists.

If not, both $s = 0$, $t = n$, it may be seen that S is positive for $\theta \in I$. From the second representation of T above it is easily seen that T, also, is positive. It follows that p satisfies the equation

$$\log_e \frac{T}{S}\frac{b_t}{b_s} = 0.$$

The left-hand side of this equation is monotone increasing in θ since

$$\min_{c,d>0} \frac{d}{d\theta}\left(\log_e \frac{T}{S}\right) > -\frac{d}{d\theta}\left(\log_e \frac{b_t}{b_s}\right) = -\frac{t-s}{\theta(1-\theta)}.$$

The existence and uniqueness of an extremum of $P(\theta)$ in I has been proved. Further, at the extremum,

$$P''(p) = \frac{Sb_s}{pq}\left[\frac{S'}{S} - \frac{T'}{T} - \frac{t-s}{pq}\right] < 0,$$

since

$$\max_{c,d}\left(\frac{S'}{S} - \frac{T'}{T}\right) = \frac{1}{p} - \frac{1}{q} < \frac{1}{pq},$$

where $q = 1-p$, we conclude that the extremum is a maximum.

In the open interval (0, 1) there is at most one local maximum. If a maximum should occur at $\theta = 0$ there could be no other local maximum, otherwise the open interval between the maxima would have to contain a minimum. The same result would hold should a maximum occur at $\theta = 1$.

In all cases there is a unique maximum of $P(\theta)$.

All that remains is to prove uniqueness of the test subject to the conditions:

$$(a) \quad P(p) = 1-\alpha,$$
$$(b) \quad P'(p) = 0,$$
$$(c) \quad P''(p) < 0.$$

Omitting, for the moment, the condition (a) on the size of the test, we find from (b) and (c) that p is an increasing function of $s^* = s+1-c$ and of $t^* = t-1+d$, independently, since

$$\frac{\partial p}{\partial s^*} = -\frac{\partial p}{\partial c} = -\frac{np-s}{P''(p)} b_s > 0,$$

$$\frac{\partial p}{\partial t^*} = \frac{\partial p}{\partial d} = -\frac{t-np}{P''(p)} b_t > 0.$$

Given p $(0 < p < 1)$, the condition (a) determines t^* uniquely as an increasing function of s^*. The value of θ which maximizes $P(\theta)$ must then be definitely increasing with s^* and cannot assume the value p more than once. It follows that the test is unique for all p in the open interval.

The special cases $p = 0$, $p = 1$, are treated differently. It is sufficient to consider the case $p = 0$. For this case there is no unique size α test. We must have $s^* = \alpha$ and t^* may take any allowable value. Unbiasedness is assured by taking $\psi_1 < 1-\alpha$. However, the uniformly most powerful test among these is given by $\psi_0 = \alpha$, $\psi_k = 0$ $(k > 0)$. A similar criterion exists for $p = 1$.

Confidence estimates

It is well known that the Neyman–Pearson method of testing may also be used as a means of estimation. The test described above leads to a confidence interval estimate of p at the confidence level $1-\alpha$. It is only necessary to tabulate s^* and t^* against p and n. To obtain the desired confidence estimate we observe the number of successes ν in n binomial trials and a value x of a random variable which has a rectangular distribution in the interval $0 \leqslant x \leqslant 1$. We obtain the respective lower and upper confidence limits p_0 and p_1 by locating in the table the values satisfying

$$t^*(p_0) = \nu - x, \quad s^*(p_1) = \nu + x.$$

REFERENCE

Tocher, K. D. (1950). Extension of the Neyman–Pearson theory of tests to discontinuous variates. *Biometrika*, **37**, 130.

A note on the circular multivariate distribution

BY G. S. WATSON

The Australian National University, Canberra, A.C.T.

1. *Summary.* In Anderson's (1941) paper on the distribution of serial correlation coefficients, Hotelling's suggestion of a circular definition of these coefficients appears. Since then this device has been used constantly to lighten the mathematical difficulties in serial correlation theory. There has, however, been no suggestion that the model would ever be a good approximation in practice—the reverse has always been suggested in fact. It is the purpose of the present note to point out a fairly common class of data in which the circular model will be a good approximation. A typical case of this occurs when one wishes to study seasonal variation and begins by averaging the data over many years.

2. *Periodically averaged data.* Let $\{x_i\}$ be generated by a stationary process with, for simplicity, zero mean and unit variance and a correlation function $\rho_s = E(x_i x_j)$, where $s = |i-j|$.

Suppose Nn consecutive x's are observed and define

$$y_j = x_j + x_{j+n} + \ldots + x_{j+(N-1)n} \quad (j = 1, \ldots, n).$$

Then, with $|i-j| = s$,

$$\operatorname{var}(y_j) = N + 2(N-1)\rho_n + 2(N-2)\rho_{2n} + \ldots + 2\rho_{(N-1)n},$$

$$\operatorname{covar}(y_i, y_j) = N\rho_s + (N-1)(\rho_{n-s}+\rho_{n+s}) + (N-2)(\rho_{2n-s}+\rho_{2n+s}) + \ldots + (\rho_{(N-1)n-s}+\rho_{(N-1)n+s})$$

and

$$\operatorname{corr.}(y_i, y_j) = \frac{\rho_s + \left(1-\frac{1}{N}\right)(\rho_{n-s}+\rho_{n+s}) + \ldots + \left(1-\frac{N-1}{N}\right)(\rho_{(N-1)n-s}+\rho_{(N-1)n+s})}{1 + 2\left(1-\frac{1}{N}\right)\rho_n + 2\left(1-\frac{2}{N}\right)\rho_{2n} + \ldots + 2\left(1-\frac{N-1}{N}\right)\rho_{(N-1)n}}.$$

When $N \to \infty$

$$\operatorname{corr.}(y_i, y_j) \to \frac{\rho_s + \rho_{n-s} + \rho_{n+s} + \ldots + \rho_{(N-1)n-s} + \rho_{(N-1)n+s}}{1 + 2\rho_n + 2\rho_{2n} + \ldots + \rho_{(N-1)n}}.$$

If in this expression s is replaced by $n-s$, the expression is only altered by the addition of $\rho_{Nn-s} - \rho_{(N-1)n+s}$ to the numerator. This additional term will usually be negligible in practice.

Thus, if N is sufficiently large, the variables y_j $(j = 1, \ldots, n)$ will have zero means, constant variances, and the correlation of y_i and y_j will be the same for $|i-j| = s$ $(s = 1, \ldots, n-1)$ as for $|i-j| = n-s$. This is the most general case of the circular multivariate distribution.

Further, if the basic process is such that ρ_r may be neglected for $r \geqslant n$, then

$$\operatorname{corr.}(y_i, y_j) = \rho_s + \rho_{n-s}, \quad |i-j| = s.$$

If, for example, the basic process is a first-order autoregressive process, with autocorrelation function $\rho_s = \rho^s$, we find that

$$\operatorname{corr.}(y_i, y_j) = \rho^s + \rho^{n-s} = \frac{\rho^s + \rho^{n-s}}{1+\rho^n},$$

since we are assuming that ρ^r $(r \geqslant n)$ is negligible. This last correlation function is that of the first-order circular autoregressive process considered by, among others, Anderson & Anderson (1950).

REFERENCES

ANDERSON, R. L. (1941). Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13**, 1.

ANDERSON, R. L. & ANDERSON, T. W. (1950). Distribution of the serial correlation coefficient for residuals from a fitted Fourier series. *Ann. Math. Statist.* **21**, 59.

The fitting of regression curves with autocorrelated data

By N. A. HUTTLY

Marconi's Wireless Telegraph Company, Chelmsford

1. Introduction

This investigation was prompted by a study of problems dealing with measurements of reflected radar signals from a moving target, one such measurement being the difference in phase between signals received at two different aerials from the same target. In this particular problem the output from the radar was a continuous wave form, so that the data under investigation was also of a continuous nature.

From physical considerations of this problem (e.g. the assumption that the target had no acceleration) it was expected that the phase difference would vary linearly with time. Thus the statistical investigation was concerned with the determination of this linear trend, that is, the problem became, in essence, that of fitting a regression line to continuously varying data. In order to deal with data of this form it is necessary to have recourse to the theory of autocorrelation, for all continuous data must be autocorrelated.

This paper is a short investigation into the general principles involved in regression fitting when the data involved is autocorrelated.

2. General discussion of the problem

In the classical theory of regression for random normal variables the method of least squares and the method of maximum likelihood are equivalent, but for stochastically dependent variables this is no longer, in general, true. In the case of large samples it has been shown that the two methods are asymptotically equivalent (Grenander, 1954; Grenander & Rosenblatt, 1954; Wold, 1950) for stationary stochastic processes. We shall, in this paper, be dealing with the problem of 'small' samples where we define a small sample as being a sample whose equivalent number of independent points is small* although the total size of the same may be large.†

We shall only be concerned here with a simplified form of stochastic dependence, namely, that in which the parameters of the process are known *a priori*. This case very often occurs in problems concerning radio signals, the process being determined by virtue of the Wiener–Khintchine relation (Rice, 1944) from the power spectrum and band width of the receiver. We have confined ourselves to the cases of linear and quadratic regression, but the methods could easily be extended to higher-order regression curves.

In §§ 3 and 4 we develop the general results appertaining to the least-squares and maximum-likelihood approaches, and in § 5 we utilize these results for a special form of stochastic dependence, the linear Markov process.

3. Least-squares solution

We may formulate as our model

$$y(t) = \alpha + \beta t + \gamma t^2 + \ldots + \epsilon(t),$$

where $y(t)$ is the observed variable; α, β, γ, the constants of the regression curve, which are to be estimated from the data; $\epsilon(t)$ is a stationary random variable.

Without loss of generality we take $\epsilon(t)$ to have zero mean, variance σ^2 and autocorrelation function $\rho(\tau)$. We also take, for convenience, the period of observation to be

$$-T \leqslant t \leqslant +T.$$

3·1. *Linear regression*

Let us consider

$$y(t) = \alpha + \beta t + \epsilon(t). \tag{1}$$

Since we sample in the range $(-T \leqslant t \leqslant +T)$ we have

$$\int_{-T}^{T} t^r \, dt = 0 \quad (r \text{ odd}). \tag{2}$$

* The sample may be discrete or continuous.

† By equivalent number of points we mean that number of independent points which gives the same amount of information as the sample considered.

In accordance with the general theory of least squares we form

$$S = \int_{-T}^{T} \{y(t) - \alpha - \beta t\}^2 \, dt,$$

and solve the equations

$$\frac{\partial S}{\partial \alpha} = \frac{\partial S}{\partial \beta} = 0.$$

This procedure leads us to the two equations

$$\hat{\alpha} = \left(\int_{-T}^{T} y(t) \, dt \right) \Big/ (2T), \tag{3}$$

$$\hat{\beta} = 3 \left(\int_{-T}^{T} t y(t) \, dt \right) \Big/ (2T^3) \tag{4}$$

(by virtue of equation (2)). Since $\hat{\alpha}, \hat{\beta}$ are functions of $y(t)$ and therefore of $\epsilon(t)$ they are random variables, therefore we can form their means and variances. From (1) we have

$$\mathscr{E}\{y(t)\} = \alpha + \beta t.$$

Therefore

$$\mathscr{E}\{\hat{\alpha}\} = \alpha, \quad \mathscr{E}\{\hat{\beta}\} = \beta,$$

and

$$\operatorname{var}(\hat{\alpha}) = \mathscr{E}\left\{ \int_{-T}^{T} \epsilon(t) \, dt \right\}^2 \Big/ (4T^2), \tag{5}$$

$$\operatorname{var}(\hat{\beta}) = 9\mathscr{E}\left\{ \int_{-T}^{T} t\epsilon(t) \, dt \right\}^2 \Big/ (4T^6). \tag{6}$$

From (5)

$$4T^2 \operatorname{var}(\hat{\alpha}) = \mathscr{E}\left\{ \int_{-T}^{T} \int_{-T}^{T} \epsilon(t) \, \epsilon(t') \, dt \, dt' \right\},$$

and on putting

$$\tau = t' - t,$$

since

$$\mathscr{E}\{\epsilon(t) \, \epsilon(t+\tau)\} = \sigma^2 \rho(\tau),$$

we have

$$4T^2 \operatorname{var}(\hat{\alpha}) = \sigma^2 \int_{-T}^{T} dt \int_{-T-t}^{T-t} \rho(\tau) \, d\tau,$$

and we reverse the order of integration so that

$$\int_{-T}^{T} dt \int_{-T-t}^{T-t} d\tau \to 2 \int_{0}^{2T} d\tau \int_{-T}^{T-\tau} dt,$$

and finally,

$$\operatorname{var}(\hat{\alpha}) = \sigma^2 \left(\int_{0}^{2T} (2T - \tau) \, \rho(\tau) \, d\tau \right) \Big/ (2T^2) \tag{7}$$

and

$$\operatorname{var}(\hat{\beta}) = 3\sigma^2 \left(\int_{0}^{2T} (4T^3 - 6T^2\tau + \tau^3) \, \rho(\tau) \, d\tau \right) \Big/ (4T^6). \tag{8}$$

3·2. *Quadratic regression*

Let us now consider

$$y(t) = \alpha + \beta t + \gamma t^2 + \epsilon(t).$$

From the normal equations we get

$$\hat{\alpha} = 3 \left(3T^2 \int_{-T}^{T} y(t) \, dt - 5 \int_{-T}^{T} t^2 y(t) \, dt \right) \Big/ (8T^3), \tag{9}$$

$$\hat{\beta} = 3 \left(\int_{-T}^{T} t y(t) \, dt \right) \Big/ 2T^3, \tag{10}$$

$$\hat{\gamma} = 15 \left(3 \int_{-T}^{T} t^2 y(t) \, dt - T^2 \int_{-T}^{T} y(t) \, dt \right) \Big/ (8T^5), \tag{11}$$

leading to

$$\mathscr{E}(\hat{\alpha}) = \alpha, \quad \mathscr{E}(\hat{\beta}) = \beta, \quad \mathscr{E}(\hat{\gamma}) = \gamma,$$

and
$$\text{var}(\hat{\alpha}) = \left\{3\sigma^2 \int_0^{2T} (48T^5 - 24T^4\tau - 80T^3\tau^2 + 60T^2\tau^3 - 5\tau^5)\,\rho(\tau)\,d\tau\right\} \Big/ (64T^6), \tag{12}$$

$$\text{var}(\hat{\beta}) = \left\{3\sigma^2 \int_0^{2T} (4T^3 - 6T^2\tau + \tau^3)\,\rho(\tau)\,d\tau\right\} \Big/ (4T^6), \tag{13}$$

$$\text{var}(\hat{\gamma}) = \left\{45\sigma^2 \int_0^{2T} (16T^5 - 40T^4\tau + 20T^2\tau^3 - 3\tau^5)\,\rho(\tau)\,d\tau\right\} \Big/ (64T^{10}). \tag{14}$$

The above analysis is true whatever the distribution of $\epsilon(t)$, but in order to effect a comparison with the maximum-likelihood solution we shall assume that the $\epsilon(t)$'s are normally distributed.

4. Maximum-likelihood solution

The method that we have used in this section is to determine the maximum-likelihood solution for the case of discrete variables, then let the distance between the variables become infinitesimal so that in the limit we have the continuous case.

Therefore we use as our model
$$y_r(t) = \alpha + \beta t_r + \gamma t_r^2 + \ldots + \epsilon_r(t).$$

4·1. *Linear regression*

We have the likelihood of obtaining a given sample of size n as
$$L = P(\epsilon_1 \ldots \epsilon_n)$$
and if the ϵ_r follow a normal distribution, then
$$L = K \exp\left[-\left(\sum_{r,s=1}^{n} \omega_{rs}\epsilon_r\epsilon_s\right) \Big/ 2\omega\right],$$
where K is a constant, ω the variance-covariance determinant of the ϵ_r's and ω_{rs} the cofactor of ρ_{rs} in ω. In order to simplify matters we make our sample of size $2n+1$ such that
$$r = -n, -(n-1), \ldots, -1, 0, 1, \ldots, (n-1), n.$$
We get, therefore, by the theory of maximum likelihood, the equations
$$\frac{\partial \log L}{\partial \alpha} = \frac{\partial \log L}{\partial \beta} = 0 \tag{15}$$
to solve.

Now since
$$\sum_{r=-n}^{n} \sum_{s=-n}^{n} \omega_{rs} t_r^q t_s^p = 0 \quad \text{if} \quad p+q \text{ is odd},$$
we get as the solutions of (15) the estimates $\bar{\alpha}$, $\bar{\beta}$ given by
$$\bar{\alpha} = (y)/(1), \tag{16}$$
$$\bar{\beta} = (ty)/(tt'), \tag{17}$$
where
$$\sum_{r=-n}^{n} \sum_{s=-n}^{n} \omega_{rs} = (1),$$
$$\sum_{r=-n}^{n} \sum_{s=-n}^{n} \omega_{rs} t_r^p t_s^q = (t^p t'^q),$$
$$\sum_{r=-n}^{n} \sum_{s=-n}^{n} \omega_{rs} y_r^p t_s^q = (y^p t^q).$$

N.B. Since $\omega_{rs} = \omega_{sr}$ (from the symmetry of the variance-covariance determinant) the above expressions are symmetrical also, i.e.
$$(t^p t'^q) = (t'^p t^q).$$

From (16) and (17) we have, using the definition of $\mathscr{E}(y_r)$,
$$\mathscr{E}(\bar{\alpha}) = \alpha, \quad \mathscr{E}(\bar{\beta}) = \beta$$
and
$$\text{var}(\bar{\alpha}) = \mathscr{E}\{(\epsilon)^2\}/(1)^2,$$
$$\text{var}(\bar{\beta}) = \mathscr{E}\{(t\epsilon)^2\}/(tt')^2.$$

But we have
$$\mathscr{E}\{(t\epsilon)^2\} = \mathscr{E}\left\{\sum_{r=-n}^{n}\sum_{s=-n}^{n}\omega_{rs}t_r\epsilon_s\right\}^2$$
$$= \mathscr{E}\left\{\sum_{r=-n}^{n}\sum_{s=-n}^{n}\sum_{u=-n}^{n}\sum_{v=-n}^{n}\omega_{rs}t_r\epsilon_s\omega_{uv}t_u\epsilon_v\right\},$$

and since
$$\mathscr{E}(\epsilon_s\epsilon_v) = \sigma^2\rho_{sv},$$
$$\mathscr{E}\{(t\epsilon)^2\} = \sigma^2\sum_{r=-n}^{n}\sum_{s=-n}^{n}\sum_{u=-n}^{n}\sum_{v=-n}^{n}\omega_{rs}\omega_{uv}t_rt_s\rho_{sv}.$$

Further, by the theory of determinants,
$$\sum_v \omega_{uv}\rho_{sv} = \begin{cases}0 & (s\neq u),\\ \omega & (s=u).\end{cases}$$

Therefore
$$\mathscr{E}\{(t\epsilon)^2\} = \omega\sigma^2\sum_{r=-n}^{n}\sum_{s=-n}^{n}\omega_{rs}t_rt_s = \omega\sigma^2(tt').$$

Hence we finally get
$$\operatorname{var}\bar{\alpha} = \frac{\sigma^2\omega}{(1)}, \tag{18}$$
$$\operatorname{var}\bar{\beta} = \sigma^2\omega/(tt'). \tag{19}$$

4·2. *Quadratic regression*

We now consider
$$y_r(t) = \alpha+\beta t_r+\gamma t_r^2+\epsilon_r(t).$$

The maximum-likelihood equations give us
$$\bar{\alpha} = \{(t^2t'^2)(y)-(t^2)(t^2y)\}/\{(t^2t'^2)(1)-(t^2)^2\}, \tag{20}$$
$$\bar{\beta} = (ty)/(tt'), \tag{21}$$
$$\bar{\gamma} = \{(1)(t^2y)-(t^2)(y)\}/\{(t^2t'^2)(1)-(t^2)^2\}, \tag{22}$$

leading to
$$\mathscr{E}(\bar{\alpha}) = \alpha, \quad \mathscr{E}(\bar{\beta}) = \beta, \quad \mathscr{E}(\bar{\gamma}) = \gamma$$

and
$$\operatorname{var}\bar{\alpha} = \omega\sigma^2(t^2t'^2)/\{(t^2t'^2)(1)-(t^2)^2\}, \tag{23}$$
$$\operatorname{var}\bar{\beta} = \omega\sigma^2/(tt'), \tag{24}$$
$$\operatorname{var}\bar{\gamma} = \omega\sigma^2(1)/\{(t^2t'^2)(1)-(t^2)^2\}. \tag{25}$$

5. Results

In §§ 3 and 4, we have derived the formal expressions for the least-squares and maximum-likelihood solutions to the problem of fitting regression curves to autocorrelated data.

By letting the summation expressions in equations (18), (19), (23), (24) and (25) tend to their integral limits we get a formal comparison between the two methods for continuous variables. In this section we shall examine the relative efficiencies of the two methods for the case of Markov dependence. This gives us
$$\rho(\tau) = e^{-m|\tau|},$$

and since m is only a scaling factor we can incorporate it into the sampling range $2T$ so that
$$\rho(\tau) = e^{-|\tau|}.$$

5·1. *Linear regression*

Substituting into (7) and (8) we get for the least-squares solution
$$\operatorname{var}(\hat{\alpha}) = \sigma^2(2T-1+e^{-2T})/2T^2, \tag{26}$$
$$\operatorname{var}(\hat{\beta}) = 3\sigma^2\{2T^3-3T^2+3-3e^{-T}(1+T)^2\}/2T^6. \tag{27}$$

For the maximum-likelihood solution we have
$$\omega = \begin{vmatrix} 1 & \rho & \rho^2 & \dots & \rho^{2n} \\ \rho & 1 & \rho & \dots & \rho^{2n-1} \\ \vdots & & & & \vdots \\ \rho^{2n} & \rho^{2n-1} & \rho^{2n-2} & \dots & 1 \end{vmatrix} = (1-\rho^2)^{2n}$$

and
$$\omega_{nn} = \omega_{-n-n} = (1-\rho^2)^{2n-1},$$
$$\omega_{rr} = (1+\rho^2)(1-\rho^2)^{2n-1},$$
$$\omega_{rr\pm1} = -\rho(1-\rho^2)^{2n-1},$$
$$\omega_{rs} = 0 \quad \text{otherwise.}$$

We sample at equal intervals throughout the range $2T$, therefore

$$t_r = \frac{rT}{n} \quad (r = -n, \ldots, -1, 0, 1, \ldots, n)$$

and
$$\rho = e^{-T/n}.$$

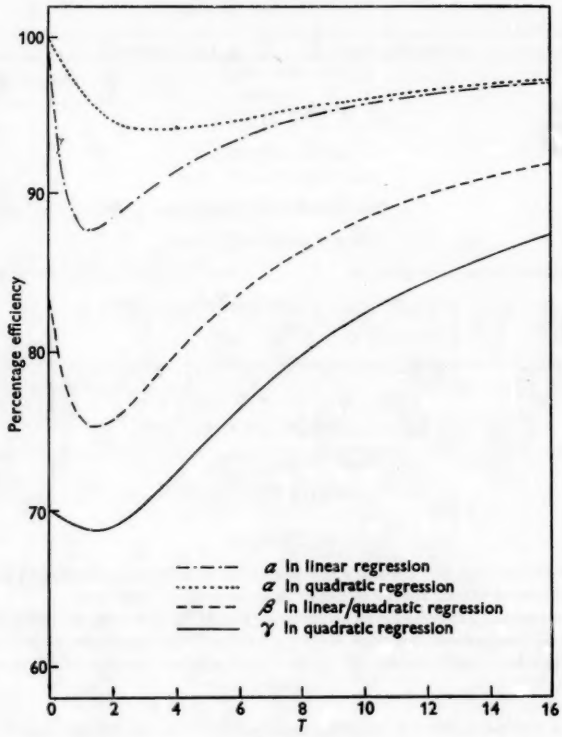


Fig. 1

This gives us
$$(1) = \sum_{-n}^{n} \omega_{rs} = (1-\rho^2)^{2n-1}\{2(1-\rho)+(1-\rho^2)^2(2n-1)\}.$$

Therefore
$$\lim_{n\to\infty} (1) = \lim_{n\to\infty} (2T/n)^{2n}(1+T),$$

and similarly for (tt'), (t^2), etc. So that upon substitution into (18) and (19) we get

$$\operatorname{var}\bar{\alpha} = \sigma^2/(T+1), \tag{28}$$
$$\operatorname{var}\bar{\beta} = 3\sigma^2/T(T^2+3T+3). \tag{29}$$

Hence by taking the ratios of (26)–(28) and (27)–(29) we can measure the efficiency of the least-squares estimators of the regression line as compared to the maximum-likelihood estimators. Fig. 1 depicts this measure as a percentage.

5·2. *Quadratic regression*

Substituting into (12), (13) and (14) we get the least-squares solution giving

$$\text{var}(\hat{\alpha}) = 3\sigma^2\{6T^5 - 3T^4 - 20T^3 + 45T^2 - 75 + 3e^{-2T}(T^2+5T+5)^2\}/(8T^6), \tag{30}$$

$$\text{var}(\hat{\beta}) = 3\sigma^2\{2T^3 - 3T^2 + 3 - 3e^{-2T}(T+1)^2\}/(2T^6), \tag{31}$$

$$\text{var}(\hat{\gamma}) = 45\sigma^2\{2T^5 - 5T^4 + 15T^2 - 45 + 5e^{-2T}(T^2+3T+3)^2\}/(8T^{10}), \tag{32}$$

and substituting into (23), (24) and (25) we get the maximum-likelihood solution

$$\text{var}(\bar{\alpha}) = 3\sigma^2(3T^2+15T+20)/\{4(T^3+6T^2+15T+15)\}, \tag{33}$$

$$\text{var}(\bar{\beta}) = 3\sigma^2/\{T(T^2+3T+3)\}, \tag{34}$$

$$\text{var}(\bar{\gamma}) = 45\sigma^2(1+T)/\{4T^3(T^3+6T^2+15T+15)\}. \tag{35}$$

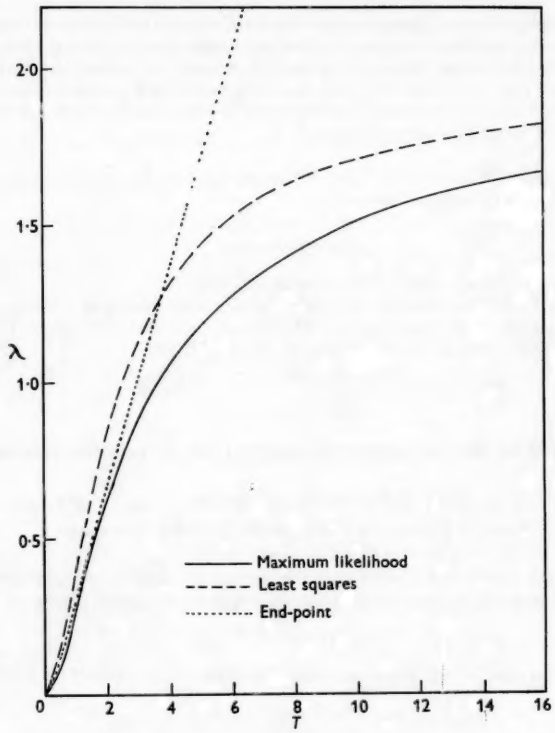


Fig. 2

Once again, the ratios of (30)–(33), (31)–(34) and (32)–(35) give us the relative efficiency of the two methods. This ratio is also depicted as a percentage on Fig. 1.

The time scale of Fig. 1 has been normalized with respect to the time scale of the parameters of the stochastic disturbance so that as plotted it is dimensionless. From Fig. 1 it is seen that the efficiency of the least-squares estimators as compared to the maximum-likelihood estimators falls to a minimum, at a time $T = 1{\cdot}5$ for the coefficients α, β of the linear regression and for β, γ of the quadratic regression, whilst it falls to a minimum at $T = 3{\cdot}5$ for the coefficient α of the quadratic regression. It is also noted that the least-squares estimator of the coefficient is more efficient in quadratic regression than in linear regression. The minimum efficiency for all cases is seen to be about 69 % for γ, showing that the method used influences the accuracy of estimation of the higher-order coefficients more than it does the lower ones.

As a corollary to this work we compare a third method of estimating the slope of a regression line. This estimator is simply obtained by taking the two end-points of the sample at hand and measuring the slope of the line joining them. We can show that for Markov dependence the variance of this estimator is

$$\sigma^2(1-e^{-2T})/2T^2.$$

Now the variances of the least-squares and maximum-likelihood estimators are $O(1/T^3)$, so we normalize them by multiplying them by $2T^3/3\sigma^2$, so that what we plot in Fig. 2 for all the three estimators is

$$\lambda = 2T^3(\text{variance of estimator})/(3\sigma^2).$$

From Fig 2 we see that the maximum-likelihood estimator is always better than the others, but for small T (<4) the end-points estimator is better than the least squares.

Conclusion

As a result of this investigation we have seen how the least-squares estimators of regression coefficients compare with maximum-likelihood estimators when autocorrelation is present. For large samples the two are asymptotically equivalent, but when the sample is small (as defined in § 2) the efficiency of the least-squares procedure can vary from 69 % for the coefficient of the quadratic term γ to 90 % for the estimate of the constant term α. So that if an efficiency of 70 % can be tolerated the method of least squares is preferable for its facility of application.

The author is indebted to Marconi's Wireless Telegraph Co. Ltd, for permission to publish this paper and to his colleagues for many helpful discussions.

REFERENCES

Grenander, U. (1954). *Ann. Math. Statist.* **25**, 252.
Grenander, U. & Rosenblatt, M. (1954). *Proc. Nat. Acad. Sci., Wash.*, **40**, 812.
Rice, S. O. (1944). *Bell Syst. Tech. J.* **23**, 282.
Wold, H. (1950). *Trans. Int. Inst. Statist.* Berne, **32**, 277.

Bounds for the variance of Kendall's rank correlation statistic

By ALAN STUART
Research Techniques Unit, London School of Economics

1. Daniels & Kendall (1947) established that the sampling variance of Kendall's rank correlation statistic (called t to distinguish it from the population parameter $\tau = E(t)$) obeys the inequality

$$V(t) \leqslant 2(1-\tau^2)/n = f_1(\tau). \tag{1}$$

They also showed that a class of rankings (called canonical rankings) existed for which

$$0{\cdot}83 \leqslant V(t)/f_1(\tau) \leqslant 1. \tag{2}$$

As they pointed out, no great sharpening of the bound to $V(t)$ can be expected in the canonical case.

In general, however, $f_1(\tau)$ is a very poor upper bound for $V(t)$, as Daniels (1950) has shown. In this paper, a new bound is proposed, which is sharper than $f_1(\tau)$ in rather more than three-quarters of all possible situations. The new bound is of the form $f_2(\tau, \rho_s)$, where ρ_s is Spearman's rank correlation (grade correlation) for the population. The ratio $f_2(\tau, \rho_s)/f_1(\tau)$ is studied, and it is found that for given τ, it decreases steadily as ρ_s does, and when ρ_s is as small as possible the ratio can be as low as $1/(n-1)$, i.e. the bound given by f_2 can be an order of magnitude in n smaller than f_1. When τ and ρ_s are equal, f_2 is about two-thirds of f_1.

2. We restrict the discussion to samples from a continuous bivariate population. In this case, Hoeffding (1947) has shown that

$$V(t) = 8\{p(1-p) + 2(n-2)(k-p^2)\}/\{n(n-1)\} \tag{3}$$

exactly, where $p = \frac{1}{2}(1+\tau)$ is the probability of concordance of a pair of bivariate observations, and k is the probability that, among three bivariate observations, a specified pair is concordant with each of the others.

We now define g as the probability that *at least* one of the three pairs of bivariate observations is concordant with each of the others, and d as the probability of complete concordance of the three pairs of observations. Then, using the formula for the probability of realization of at least one among compatible events, we have

$$g = 3k - 2d. \tag{4}$$

Hoeffding (1948) has shown that

$$g = \tfrac{1}{2}(1+\rho_s), \tag{5}$$

where ρ_s is the population grade correlation, the analogue of Spearman's rank correlation statistic r_s. Using (4) and (5) in (3), we find

$$V(t) = \frac{2}{n(n-1)}\{1-\tau^2+\tfrac{3}{2}(n-2)[2(1+\rho_s+4d)-3(1+\tau)^2]\}. \tag{6}$$

An unbiased estimator of $V(t)$ may be constructed from (6) which is simpler to calculate than the equivalent form based on (3) or that suggested by Daniels & Kendall (1947). For n at all large, it is the tediousness of estimating d in (6) or k in (3), each of which requires examination of all $\binom{n}{3}$ triplets of observations in the sample, which leads us to seek a bound for $V(t)$ instead.

3. Mann (1945) gives a simple proof that

$$\text{Prob}\{X_1 < X_2 < X_3\} \leqslant \text{Prob}\{X_1 < X_2\}\,\text{Prob}\{X_2 < X_3\}. \tag{7}$$

Applied here, (7) gives at once

$$d \leqslant p^2 = \tfrac{1}{4}(1+\tau)^2, \tag{8}$$

and use of (8) in (6) gives us our new bound

$$V(t) \leqslant \frac{2}{3n(n-1)}\{(2n-1)(1-\tau^2)+4(n-2)(\rho_s-\tau)\} = f_2(\tau,\rho_s). \tag{9}$$

Since

$$\tau^2 = E(t^2) - V(t)$$

and (Hoeffding, 1948)

$$E(t) = \tau, \quad E(r_s) = \{3\tau+(n-2)\rho_s\}/(n+1),$$

we find from (9) a bound for an unbiased estimator, $\hat{V}(t)$, of $V(t)$, which is

$$\hat{V}(t) \leqslant \frac{2}{(3n-1)(n-2)}\{(2n-1)(1-t^2)+4(n+1)(r_s-t)\}. \tag{9a}$$

The inequality in (9) becomes a strict equality only when that in (8) does, i.e. when $\tau = \rho = +1$ or -1. In this degenerate case $V(t) = f_2(\tau,\rho_s) = 0$, from (9).

4. The difference between the two bounds is

$$f_1(\tau) - f_2(\tau,\rho_s) = \frac{2(n-2)}{3n(n-1)}\{1-\tau^2-4(\rho_s-\tau)\}, \tag{10}$$

and is of order n^{-1}, just as the bounds themselves are. From (10) we deduce that

$$f_1(\tau) > f_2(\tau,\rho_s),$$

if and only if

$$\rho_s < \tfrac{1}{4}(1+4\tau-\tau^2). \tag{11}$$

If $\tau = \rho_s = +1$ or $= -1$, the bounds are both zero, as is the variance.

5. We now study the behaviour of the ratio of the bounds in the interval $-1 < \tau < 1$, namely,

$$f_2(\tau,\rho_s)/f_1(\tau) = \frac{2n-1}{3(n-1)} + \frac{4(n-2)(\rho_s-\tau)}{3(n-1)(1-\tau^2)}. \tag{12}$$

The only part of (12) we need consider is

$$F(\tau,\rho_s)=\frac{(\rho_s-\tau)}{1-\tau^2}. \tag{13}$$

We approach the problem by considering ρ_s as a function of τ, and asking what form this function must take to give $F(\tau,\rho_s)$ stationary values. Differentiating with respect to τ, we have

$$F'(\tau,\rho_s)=\frac{(\rho_s'-1)(1-\tau^2)+2\tau(\rho_s-\tau)}{(1-\tau^2)^2},$$

and this is zero when

$$\rho_s'=\frac{2\tau\rho_s-(1+\tau^2)}{\tau^2-1}. \tag{14}$$

This may be solved as a differential equation in τ, and using the initial condition $\rho_s(1)=1$, we obtain as the unique solution

$$\rho_s=\tau. \tag{15}$$

Substitution of (15) into (12) gives

$$f_2(\tau,\tau)/f_1(\tau)=\tfrac{1}{3}(2n-1)/(n-1) \tag{16}$$

independent of the value of τ. From (12) and (16), it follows that $f_2(\tau,\rho_s)/f_1(\tau)-\frac{1}{3}(2n-1)/(n-1)$ will be positive or negative with $(\rho_s-\tau)$. Thus (15) represents a 'shoulder' in the surface of $F(\tau,\rho_s)$. There is no turning point.

6. For fixed τ, (12) is a monotone increasing function of ρ_s. It is thus as small as possible when ρ_s is a minimum. Daniels (1950) and Durbin & Stuart (1951) have established the sharp bounds for ρ_s:

$$\left.\begin{aligned}\tfrac{1}{2}(3\tau-1)\leqslant\rho_s\leqslant\tfrac{1}{2}(1+2\tau-\tau^2)\quad&\text{for}\quad\tau>0,\\ \tfrac{1}{2}(\tau^2+2\tau-1)\leqslant\rho_s\leqslant\tfrac{1}{2}(3\tau+1)\quad&\text{for}\quad\tau\leqslant 0.\end{aligned}\right\} \tag{17}$$

Substituting into (12) the lower bounds for ρ_s in (17), we obtain

$$\left.\begin{aligned}f_2\{\tau,\tfrac{1}{2}(3\tau-1)\}/f_1(\tau)&=\frac{(2n-1)\tau+3}{3(n-1)(1+\tau)}\qquad(1>\tau>0),\\ f_2\{\tau,\tfrac{1}{2}(\tau^2+2\tau-1)\}/f_1(\tau)&=\frac{1}{n-1}\qquad(-1<\tau\leqslant 0).\end{aligned}\right\} \tag{18}$$

The ratios in (18) range from $\frac{1}{3}(n+1)/(n-1)$ near $\tau=1$ to $1/(n-1)$ for $\tau\leqslant 0$. It is remarkable that we have from (9) the bound

$$f_2\{\tau,\tfrac{1}{2}(\tau^2+2\tau-1)\}=\frac{2}{n(n-1)}(1-\tau^2)\quad(-1<\tau\leqslant 0), \tag{19}$$

of order n^{-2}. (This is not a new type of result; Hoeffding ((1948), p. 318) gave an example where

$$V(t)=\frac{2}{n(n-1)}$$

exactly.) Thus the bound (9) is an order of magnitude in n better than (1) in extreme cases.

7. In the canonical ranking situations discussed by Daniels & Kendall (1947), it is easy to show that

$$\rho_s\doteqdot 1-2\left(\frac{1-\tau}{2}\right)^{\frac{1}{2}}. \tag{20}$$

In (20) $\tau\geqslant 0$ applies to the inverse canonical ranking, while $\tau\leqslant 0$ applies to the direct canonical ranking. Substitution of (20) into (12) gives

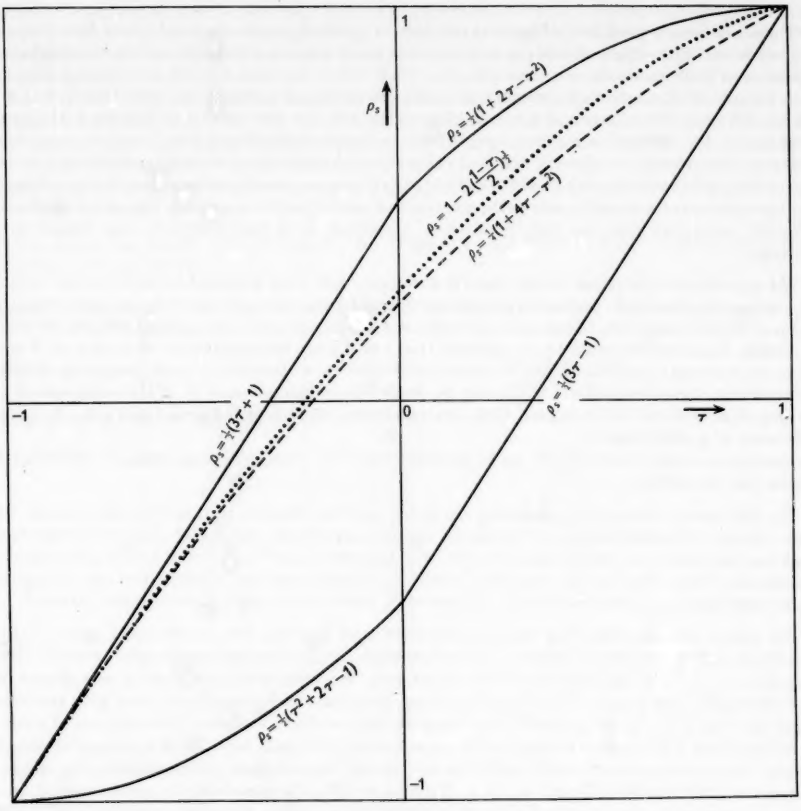
$$f_2\left\{\tau,1-2\left(\frac{1+|\tau|}{2}\right)^{\frac{1}{2}}\right\}\Big/f_1(\tau)\doteqdot 1+\frac{4(n-2)}{3(n-1)}\left\{\frac{1-[\frac{1}{2}(1-\tau)]^{\frac{1}{2}}}{1+\tau}-\frac{1}{4}\right\}. \tag{21}$$

The expression in braces on the right of (21) is

$$\frac{1-[\frac{1}{2}(1-\tau)]^{\frac{1}{2}}}{1+\tau}-\tfrac{1}{4}>0\quad(-1<\tau<1), \tag{22}$$

so that (21) is >1, and therefore f_1 always provides a better bound than f_2 in the canonical case.

8. The diagram illustrates our results. The full lines delimit the possible combinations of the coefficients: f_2 is a better bound than f_1 for all points below the dashed line. The dotted line, representing the canonical situations, lies entirely above the dashed line. Integrations show that f_2 is a better bound than f_1 in $\frac{11}{14}$, or 78·6 %, of the possible situations.



REFERENCES

DANIELS, H. E. (1950). Rank correlation and population models. *J. R. Statist. Soc.* B, **12**, 171.

DANIELS, H. E. & KENDALL, M. G. (1947). The significance of rank correlations when parental correlation exists. *Biometrika*, **34**, 197.

DURBIN, J. & STUART, A. (1951). Inversions and rank correlation coefficients. *J. R. Statist. Soc.* B, **13**, 303.

HOEFFDING, WASSILY (1947). On the distribution of the rank correlation coefficient τ when the variates are not independent. *Biometrika*, **34**, 183.

HOEFFDING, WASSILY (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293.

MANN, HENRY B. (1945). Non-parametric tests against trend. *Econometrica*, **13**, 245.

A note on the theory of quick tests

By D. R. COX

Department of Biostatistics, School of Public Health, University of North Carolina

1. There has been a good deal of interest recently in quick approximate methods of examining statistical significance. The object of this paper is to make some general comments on the interpretation and justification of such methods. An example of a quick test is the sign test for a normal mean; given a random sample of size n from a normal population of unknown mean μ, the hypothesis that $\mu = 0$ is tested by referring the number of positive observations to the binomial distribution with parameter $\frac{1}{2}$ and index n. The efficient test, the t test, is here so simple to work out that it is not very often that a quick test is required to replace it, the real value of quick tests being in more complicated situations, where the fully efficient test may be difficult to apply. Thus in a problem where the efficient analysis calls for say the solution of a complex non-orthogonal set of least-squares equations, the use of an alternative 'inefficient' procedure may be very profitable. However, it is convenient to use simple cases for illustration.

2. We consider in this paper a very special situation. Let it be required to test the null hypothesis $\theta = \theta_0$ concerning the single unknown parameter θ, and let the theoretical set-up be sufficiently simple for the best significance test, based on a statistic t say, to be uniquely determined and let the proposed quick test be based on the statistic, q. Assume that t and q are both unbiased estimates of θ and deal with the large-sample approximation in which t and q follow a bivariate normal frequency distribution with a variance-covariance matrix which can be treated as independent of θ. If the variances of t and q are σ^2 and σ^2/E, Fisher (1925) showed that the correlation coefficient between t and q is $\sqrt{E}$, where E is the efficiency of q relative to t.

The general remarks below apply quite generally but the quantitative analysis is restricted to the situations just described.

3. The customary method of comparing the quick and the efficient tests is by power curves. That is, we plot against θ the probability of attaining significance at say the α per cent level, first with the efficient test and then with the quick test. The drop in power from the efficient to the quick test is then an indication of the effect on the long-run behaviour of replacing the efficient test by the quick test. In the present case the power curves are expressed in terms of the normal probability integral.

4. The power curves define the long-run behaviour of the two tests considered individually, but give no direct indication of the extent to which the answers obtained by applying the tests to the same data tend to agree. This last, however, seems of interest. Some knowledge of it gives us a clearer picture of the comparative behaviour of the two tests in applications, and also helps to deal with the following situation that can arise in using quick tests. Suppose that we have made a quick analysis of some data. If the full analysis would take an appreciable time we may well ask: how likely is it that we should get essentially the same answer from the efficient test as we have already obtained from the quick test? If we are convinced that the efficient analysis, if made, would only reproduce the conclusions of the quick analysis, we may rest content with the quick analysis. If, however, there is a possibility of materially different conclusions, i.e. of a greatly changed level of significance, we would often decide to go ahead with the full analysis. This would be especially so in scientific work where the data are to be published and may be reanalysed by future workers.

5. The considerations of § 4 suggest that we should examine the regression of t on q, i.e. that we should consider the problem of predicting the efficient statistic t given the value of the quick statistic, q. An alternative formulation would be to ask how likely it is, given q, that t exceeds its α per cent critical value for testing the null hypothesis. This approach seems, however, rather less helpful since we are usually interested to some extent in the actual level of significance attained by the data and not just with whether some preselected level is exceeded.

6. For given θ, q it follows from the properties of the bivariate normal surface that t/σ is normally distributed with mean $\theta(1-E)/\sigma + qE/\sigma$ and variance $1-E$. This involves the unknown parameter θ and while we could, for some purposes, consider the whole set of joint distributions as a function of the unknown parameter θ, it is an advantage to introduce an averaging operation over θ. Therefore we assume Bayes's hypothesis that the prior distribution of θ is uniform; then, since, for given θ, q is normally distributed around θ with variance σ^2/E, θ, for given q, is normally distributed around q with variance σ^2/E. Hence t/σ is normally distributed with mean q/σ and variance $(1-E)/E$. Or equivalently, if the

normal deviate based on q for testing the null hypothesis $\theta = \theta_0$ is $q' = (q-\theta_0)\sqrt{E}/\sigma$, the corresponding normal deviate, $t' = (t-\theta_0)/\sigma$ has, given q', mean $q'/\sqrt{E}$ and variance $(1-E)/E$.

7. The distribution of t' given q' just derived depends for its frequency interpretation on the truth of Bayes's hypothesis as a statement about the frequencies of values of θ in a long series of applications of the test. It is not usual to base a theory of statistical inference on Bayes's hypothesis without abandoning the frequency interpretation of probability, but in the present paper we are merely exploring in a general way the properties of quick tests. It is perfectly reasonable to ask how the quick test would behave if θ were distributed in some special way, and if a special form is to be chosen the uniform distribution seems a natural one.

The distribution of t' given q' in the range of most practical interest depends on the form of the prior distribution in the interval, say $|\theta-\theta_0| < 6\sigma$; in large samples σ will be small and so this will represent a narrow interval for θ over which the prior probability is, in some types of application, likely to vary little. If a prior distribution for θ exists and is not constant over this range the results will be affected. For example, if the prior probability density is higher at the centre of the interval, the value of $|t'|$ given q' will be reduced.

8. A few numerical results derived from the formulae of § 6 are given in Table 1. Some examples will now be given to illustrate this table.

Table 1. *Relation between standardized deviates t', q' when θ has a special prior distribution*

E	Mean of t' given q'	Standard error of t' given q'
0·99	1·005	0·101
0·95	1·026	0·229
0·90	1·054	0·333
0·80	1·118	0·500
0·70	1·195	0·655
0·60	1·291	0·816
0·50	1·414	1·000
0·40	1·581	1·225
0·30	1·826	1·528
0·20	2·236	2·000
0·10	3·162	3·000

Example 1. Suppose that a test with $E = 0{\cdot}99$ gives a value just significant at 5%, i.e. $q' = 1{\cdot}645$. Then t' has mean 1·653 and standard error 0·101. Therefore t' probably lies between

$$1{\cdot}653 \pm 0{\cdot}101 = (1{\cdot}552, 1{\cdot}754)$$

and fairly certainly lies between

$$1{\cdot}653 \pm 1{\cdot}96 \times 0{\cdot}101 = (1{\cdot}455, 1{\cdot}851).$$

The significance levels corresponding to the two sets of limits are approximately (0·060, 0·040) and (0·073, 0·032).

As would be expected with so high an efficiency, serious disagreement between the results of the tests is unlikely.

Example 2. Suppose that a test with $E = 0{\cdot}80$ gives satisfactory agreement with the null hypothesis, say $q' = \frac{1}{2}$. Then t' has mean 0·559 and standard error $\frac{1}{2}$. The two sets of limits for t' are therefore (0·059, 1·059) and $(-0{\cdot}421, 1{\cdot}539)$. Therefore it is unlikely that the level of significance with t' is higher than about 6%.

For many purposes the efficiency would be considered quite high, and yet appreciable disagreement between the two tests will arise quite often.

Example 3. Consider a very inefficient test with $E = 0{\cdot}2$ and suppose that it gives exact agreement with the null hypothesis, i.e. $q' = 0$. Then since the standard error of t' given q' is 2, there is an appreciable chance that t' would have given significance at a high level. If, on the other hand, $q' = 2$, indicating significance at the $2\frac{1}{2}$% level, t' has mean 4·472 and standard error 2, so that it is rather unlikely that t' fails to indicate high significance.

This example expresses quantitatively the obvious point that failure to get significance with a very inefficient test gives no information about whether significance would be obtained with an efficient test, but that if the quick test gives significance it is likely that the efficient one will too.

It must again be stressed that these are illustrations of the relation between t' and q' for a special prior distribution of θ.

9. The estimation formulae analogous to the above results for significance tests are as follows. Let q be an estimate of θ with standard error σ_q and efficiency E, and let t be the efficient estimate from the same sample. Then if θ has a uniform prior distribution, $(t-q)$ has mean 0 and standard error $\sigma_q\sqrt{(1-E)}$, thus enabling us to assess from the quick estimate whether our conclusions are likely to be appreciably modified by computing the efficient estimate. This will not be discussed further here.

10. Two general points need to be made about the above calculations. First, it is in practice sometimes possible to estimate approximately the probable direction and nature of the difference between t and q by careful inspection of the observations, the main difference between t and q often being in the weight they attach to extreme observations. Secondly, all the above work is based on what happens when the observations follow the distribution law assumed in deriving the tests; the relative robustness of the two tests, which is relevant in choosing between them, has not been considered.

11. The general purpose of these results has been discussed in § 4. A direct application of the formulae to the analysis of data would be as follows: First apply the quick test; then by the method of § 7 work out the probable limits for t'—to do this the efficiency E of the quick method needs to be known. If it seems possible that t' may give a seriously different conclusion from q', t' itself is calculated and the significance level given by it used. Of course the selection of that test giving the more significant answer is wrong. This procedure needs further investigation before it can be generally recommended.

Example 4. In a test of two methods A, B of determining by sampling the total number of bacteria on a plate, twenty plates were counted by both methods. The differences between the counts by A and B were 3, −3, 3, 6, −8, −1, 6, 8, 5, −8, 11, −1, 5, 6, 5, −6, 3, 7, 14, 8. Is there evidence of bias?

There are 14 positive observations out of 20 so that, using the sign test, $q' = (14-10)\sqrt{4}/\sqrt{20} = 1{\cdot}79$, which is significant at 10 % in a two-sided test. (If we correct for continuity the result is less significant.) Since for this test $E = 2/\pi$ (Cochran, 1937), t' has mean $1{\cdot}79\sqrt{\pi}/\sqrt{2}$ and standard error $\sqrt{[(\pi-2)/\pi]}$, i.e. is $2{\cdot}24 \pm 0{\cdot}60$, so that t', which in this case is Student's t, could, consistently with q', take values representing widely differing significance levels. Therefore it is worth working out t', and its value comes to $2{\cdot}35\,(0{\cdot}05 > P > 0{\cdot}02)$, in good agreement with the value predicted from q'. Notice that there are no very large deviations; if say the second observation had been −18, instead of −3, q' would have been unaltered. However, we would have noted this large negative deviation and would not have expected t' to give a more significant answer than the quick test.

Summary. It is suggested that in comparing an inefficient quick test with an efficient test we should consider the extent to which the results of applying the two tests to the same data tend to agree, in addition to the long-run behaviour of the tests considered separately. A very simple case is considered quantitatively.

I am grateful to Prof. E. S. Pearson for very helpful criticism.

REFERENCES

COCHRAN, W. G. (1937). The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient. *J. R. Statist. Soc.* **100**, 69–73.

FISHER, R. A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.* **22**, 700–25.

A note on the signs of gross correlation coefficients and partial correlation coefficients

BY OLAV REIERSØL

University of Oslo, Norway

We shall use the following matrix theorem:

THEOREM 1. *If the principal minor determinants of a square matrix A and the determinant value of A itself are all positive, and if the non-diagonal elements of A are all negative, then all elements of the adjoint of A are positive.*

This theorem was proved in a slightly different form by Mosak (1944, pp. 49–51). The theorem was stated in its present form by Metzler (1950, p. 340).

Remark. Theorem 1 still remains true if 'adjoint' is replaced by 'inverse'.

It is evident from the proof of Metzler that Theorem 1 may be sharpened to

THEOREM 1′. *If the principal minor determinants of a square matrix A are all non-negative, and if the non-diagonal elements of A are all negative, then all non-diagonal elements of the adjoint of A are positive.*

Let

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix}$$

be the correlation matrix of a set of variables $x_1, x_2, \dots, x_n$. The partial correlation coefficients of highest order of the set are given by

$$p_{ij} = -\frac{\text{cof.}\,(r_{ij})}{\sqrt{\{\text{cof.}\,(r_{ii})\,\text{cof.}\,(r_{jj})\}}}, \tag{1}$$

where cof. (r_{ij}) is the cofactor of r_{ij} in the matrix R. We see that the sign of a partial correlation coefficient is the opposite of the sign of the corresponding element of the adjoint of R. Applying Theorem 1′ we thus get:

THEOREM 2. *If all gross correlation coefficients of a set of variables are negative, all partial correlation coefficients of all orders of the set are also negative.*

We shall next consider the case when all partial correlation coefficients of highest order are positive. Then all non-diagonal elements of the adjoint of R are negative.

Suppose first that R is singular. Then the rows of the adjoint of R are proportional. Since the diagonal elements of the adjoint of R are all positive, the non-diagonal elements cannot all be negative when $n \geqslant 3$. We conclude that R is non-singular when all partial correlation coefficients of highest order are positive. The inverse matrix $Q = R^{-1}$ thus exists, and its non-diagonal elements are all negative. Applying Theorem 1 we conclude that the elements of $R = Q^{-1}$ are all positive.

Let R and Q be partitioned symmetrically into submatrices in the same way

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

Then (Hotelling, 1943, p. 4)

$$R_{11}^{-1} = Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}.$$

Using the information we have about the signs of the submatrices of Q and applying Theorem 1 to the matrix Q_{22}, we conclude that the non-diagonal elements of R_{11}^{-1} are all negative. Now any partial correlation coefficient of any order of the set $x_1, x_2, \dots, x_n$ has opposite sign of a non-diagonal element of some R_{11}^{-1} of the original correlation matrix or of a matrix which we obtain from R by an interchange of rows and the corresponding interchange of columns, hence we obtain:

THEOREM 3. *If all partial correlation coefficients of highest order of a set of variables are positive, all partial correlation coefficients of lower orders are also positive, and all gross correlation coefficients are positive.*

If we change the signs of all elements of one or more of the rows of a square matrix A and afterwards change the signs of all elements of the corresponding columns, we have performed a particular kind of cogredient (also called congruent) transformation of the matrix A. Moreover, this is the only cogredient transformation whose only effect is to change signs. We shall therefore call it a *cogredient change of signs* of the matrix A. We shall adopt the convention that no change of signs is also a cogredient change of signs.

A cogredient change of signs of the matrix R will also be called a cogredient change of signs of the gross correlation coefficients of the set $x_1, \dots, x_n$. If P is the matrix whose elements are defined by (1), a cogredient change of signs of P will also be called a cogredient change of signs of the partial correlation coefficients of highest order of the set of variables $x_1, \dots, x_n$.

The following generalization of Theorems 2 and 3 is easily proved by performing a change of signs of a subset of the set of variables $x_1, \dots, x_n$:

THEOREM 4. *If the signs of all gross correlation coefficients of a set of variables may be made negative by a cogredient change of signs, or if the signs of all partial correlation coefficients of highest order of the set may be made positive by a cogredient change of signs, then, without any change of signs, any partial correlation coefficient of any order has the same sign as the corresponding gross correlation coefficient.*

REFERENCES

HOTELLING, H. (1943). Some new methods in matrix calculation. *Ann. Math. Statist.* **14**, 1–34.
METZLER, L. A. (1950). A multiple-region theory of income and trade. *Econometrica*, **18**, 329–54.
MOSAK, J. L. (1944). *General Equilibrium Theory of International Trade*. Cowles Commission Monograph no. 7, Bloomington, Indiana.

The estimation of the mean of a censored normal distribution by ordered variables

BY P. G. MOORE
University College, London

1. In many practical problems the final data obtained are incomplete in that time or equipment has not been available for completion of the originally planned experiment or that something has gone wrong, leaving some observations missing. We suppose that the observations form a random sample from a normal population, but that the sample is in some way censored, meaning that all observations above some limit, or below another limit or outside both limits, are unavailable to us for the purpose of estimating the mean. The only information available to us about the missing observations is their total number.

2. One method of approach is to consider the maximum-likelihood solution. Details of this with appropriate auxiliary tables have been given by Hald (1949) and Gupta (1952). In the case of small samples the two solutions differ. Hald supposes that of the n observations in the sample only the r values of the variable, x, which fall on one side of a fixed truncation point are available; r will thus vary binomially from sample to sample. Gupta supposes that if the observations arranged in order of magnitude are $x_1, x_2, \ldots, x_n$, then the values of $x_1 \ldots x_r$ are known but not the values of $x_{r+1} \ldots x_n$. Here r is fixed. The asymptotic variance of both estimators is, however, the same, and it is this variance which is used below in comparing the efficiency of the maximum-likelihood estimator with alternative estimators.

The purpose of this note is to point out that by using methods of estimation of the mean suggested in an editorial paper of Karl Pearson (1920) we can often get results that are nearly as good as the maximum-likelihood results at a great saving of time and calculation. Pearson's technique depends on the fact that even though a sample has been censored many of the ordered variables are available to us and the problem is how best to use those values.

3. For computations use will be made of the expansions given by David & Johnson (1954) for ordered variables in a random sample of n individuals drawn from a population whose probability density function is $f(x)$ with mean ξ and standard deviation unity. We define

$$F(X) = \int_{-\infty}^{X} f(x)\,dx, \tag{1}$$

and arrange the sample observations in order of magnitude to give values $x_1, x_2, \ldots, x_n$. Defining X_r by the relation

$$F(X_r) = \frac{r}{n+1} \quad (r = 1, 2, \ldots, n)$$

and expanding x_r about X_r in an inverse Taylor series gives

$$x_r = X_r + X_r'[h(x_r)] + \tfrac{1}{2}X_r''[h(x_r)]^2 + \ldots, \tag{2}$$

where $$h(x_r) = F(x_r) - F(X_r) \quad \text{and} \quad X_r' = \frac{dX_r}{dF(X_r)} \quad \text{and so on.}$$

The utilization of such expansions enables various approximations to the moments of ordered variables to be obtained, and David & Johnson give tables of them in terms of powers of $(n+2)^{-1}$. The accuracy of these expansions was examined for a few cases of sampling from a normal population. For example, using (2), the variance of the median is

$$\text{var}\,(x_m) = \frac{1{\cdot}570796}{n+2} + \frac{2{\cdot}467401}{(n+2)^2} + \frac{3{\cdot}462732}{(n+2)^3}, \tag{3}$$

and numerical values derived from this formula are compared below with exact values for low sample sizes given by Hoie (1931):

Median in sample of n	$n=5$	$n=7$	$n=9$	$n=11$
Exact variance	0·286834	0·210446	0·166093	0·137227
Approximate variance	0·284849	0·209744	0·165793	0·137164

For samples up to size 10 the variances and covariances of all ordered variables have been given by Godwin (1949). As an illustration, we compare two cases for samples of $n = 10$ using terms in (2) up to $(n+2)^{-3}$. From these examples it seems that the expansions would give sufficient accuracy for practical purposes provided n was greater than about 10.

	var (x_3)	covar $(x_3 x_8)$
Exact value	0·17500	0·06302
Approximate value	0·17485	0·06297

4. If we use a pair of the ordered variates to estimate the mean, there are two cases to be considered, as we may either use two symmetrically placed variates, that is, x_r and x_{n-r+1}, or we may alternatively use an unsymmetrical pair. For a one-sided censored form of distribution more observations are available to us on one side of the central value than on the other, and to make use of a symmetrical pair means that a number of observations may appear to be needlessly wasted. Considering the symmetrical case first we find from the expressions in (2) that for large samples the estimate

$$V_2 = \tfrac{1}{2}(x_r + x_{n+1-r}) \tag{4}$$

has as variance the approximate value $\text{variance}\,(V_2) \doteqdot \dfrac{X_r'^2 p_r}{(n+2)}$, (5)

where $p_r = F(X_r) = r/(n+1)$.

We find, on substitution, that the asymptotic efficiency of V_2 as compared with that of the mean of the complete sample is $2z_r^2/p_r$, where z_r is the ordinate of the standardized normal distribution corresponding to a probability integral of p_r. This function is tabulated in Table 1 together with the asymptotic maximum-likelihood efficiencies from Gupta's work based on the assumption that the distribution is censored at one end only. From the table we see that the efficiency of V_2 increases with p_r to a maximum at about p_r equal to 0·27 and then decreases to the value 0·637 which occurs when just the median is used for the estimation. Pearson showed that the maximum efficiency occurs when p_r is equal to 0·271, and the table shows that if the censoring at either end is greater than 27·1 % we should take the symmetrical pair that lies farthest out from the centre of the distribution, whilst if it is less we should take the pair that makes $r/(n+1)$ as near 0·27 as possible.

Table 1. *Efficiencies of estimates using a pair of quantiles*

p_r	Efficiency of V_2	Maximum-likelihood efficiency	p_r	Efficiency of V_2	Maximum-likelihood efficiency	p_r	Efficiency of V_2	Maximum-likelihood efficiency
0·10	0·616	0·980	0·24	0·805	0·920	0·38	0·763	0·806
0·12	0·667	0·974	0·26	0·809	0·907	0·40	0·746	0·784
0·14	0·708	0·967	0·28	0·809	0·893	0·42	0·728	0·763
0·16	0·740	0·959	0·30	0·806	0·878	0·44	0·707	0·738
0·18	0·765	0·950	0·32	0·799	0·861	0·46	0·685	0·714
0·20	0·784	0·941	0·34	0·790	0·845	0·48	0·661	0·685
0·22	0·797	0·930	0·36	0·778	0·826	0·50	0·637	0·659

Next we consider whether it is possible to improve the estimates by using an unsymmetrical pair of quantiles. In this case we have to use a weighted average of the pair. Let the two variates be x_r and x_s and consider

$$\hat{\xi} = \frac{x_r \Phi\left(\dfrac{s}{n+1}\right) - x_s \Phi\left(\dfrac{r}{n+1}\right)}{\Phi\left(\dfrac{s}{n+1}\right) - \Phi\left(\dfrac{r}{n+1}\right)} = \bar{V}_2 \quad \text{(say)}, \tag{6}$$

where $\Phi(V)$ is the deviate of the standardized normal curve corresponding to a cumulative probability of V. Then if the unknown underlying normal distribution has mean ξ and standard deviation σ, we have

$$\mathscr{E}\left(\frac{x_r-\xi}{\sigma}\right) = \Phi\left(\frac{r}{n+1}\right), \quad \mathscr{E}\left(\frac{x_s-\xi}{\sigma}\right) = \Phi\left(\frac{s}{n+1}\right),$$

and hence on substitution in (6) and reduction, we have

$$\mathscr{E}(\hat{\xi}) = \xi.$$

Further, as an approximation to the variance, we find

$$(n+2)\operatorname{var}(\bar{V}_2) \doteqdot \alpha^2 \frac{p_r(1-p_r)}{z_r^2} + (1-\alpha)^2 \frac{p_s(1-p_s)}{z_s^2} + 2\alpha(1-\alpha)\frac{p_r(1-p_s)}{z_r z_s}, \tag{7}$$

where

$$\alpha = \Phi\left(\frac{s}{n+1}\right) \Big/ \left\{\Phi\left(\frac{s}{n+1}\right) - \Phi\left(\frac{r}{n+1}\right)\right\}.$$

Using (6) and (7), the optimum values of p_s have been found for four different values of p_r together with the new efficiencies. Although there is some gain, the gain is not in general very large.

Table 2. *Efficiencies of* $\bar{V}_2$

	p_r	0·30	0·35	0·40	0·45
Symmetrical	p_s	0·70	0·65	0·60	0·55
	Efficiency	0·806	0·784	0·746	0·696
Unsymmetrical	p_s	0·74	0·73	0·72	0·67
	Efficiency	0·808	0·794	0·764	0·716

5. To improve the efficiency of the estimate four ordered variates placed in two pairs could be used. Thus the estimator

$$V_4 = \tfrac{1}{2}\alpha(x_r + x_{n+1-r}) + \tfrac{1}{2}(1-\alpha)(x_s + x_{n+1-s})$$

would have the mean of the sampled population as its expected value. The efficiency would vary with the value of α that is taken and also with the values of r and s chosen. We will assume that $0{\cdot}5 \geqslant p_s > p_r$. Table 3 is a modified form of Pearson's table and shows, for large samples, the optimum values of p_s

Table 3. *Efficiencies of* V_4

p_r	0·1	0·2	0·25
p_s	0·35	0·42	0·42
α	0·374	0·542	0·615
Efficiency	0·912	0·897	0·872
Maximum-likelihood efficiency	0·980	0·941	0·913

Table 4. *Efficiencies of* V_4 *with* $p_s = 0{\cdot}5$

p_r	0·1	0·2	0·3	0·4
α	0·51	0·38	0·25	0·15
Efficiency	0·866	0·878	0·834	0·750
Maximum-likelihood efficiency	0·980	0·941	0·878	0·784

and α for various values of p_r together with the efficiency thereby achieved and the corresponding maximum-likelihood efficiency when a portion p_r is censored. In Table 4 we take the case where p_s is equal to 0·5, i.e. we use the median and two symmetrically placed variates. The optimum efficiencies are now somewhat greater than when just two variates were used and are approaching the maximum-likelihood values.

6. The estimates could be improved by using more variates but would then become rather unwieldy. The results obtained would seem to show that if there is a fair amount of censoring, say over 20 % of the distribution, then the simple estimates based on ordered variables may suffer very little in efficiency when compared with a maximum-likelihood estimate of the mean. On the other hand, when there is very little censoring, say under 10 %, there is a considerable gain in using the maximum-likelihood methods of estimation. In a situation which involves censoring at both ends of the distribution, the ordered variates would be even better *vis-à-vis* the one-ended maximum-likelihood estimate, but such cases would occur rarely in practice.

Later note. The recent publication of a paper by D. Teichroew (1956, *Ann. Math. Statist.* **27**, 410) enables the exact efficiency of V_2 to be calculated and compared with the approximate probabilities obtained as in Table 1 for the case of a sample of size 20.

r	p_r	Efficiency of V_2	
		Exact	Approximate
4	0·1905	0·780	0·775
5	0·2381	0·816	0·805
6	0·2857	0·824	0·808
7	0·3333	0·810	0·796
8	0·3810	0·779	0·763

Thus the approximate formula slightly underestimates the efficiencies for a sample of size 20 and V_2 is in fact slightly better, as compared with the maximum-likelihood estimator, than it appeared to be from Table 1.

REFERENCES

David, F. N. & Johnson, N. L. (1954). *Biometrika*, **41**, 228.
Godwin, H. J. (1949). *Ann. Math. Statist.* **20**, 279.
Gupta, A. K. (1952). *Biometrika*, **39**, 260.
Hald, A. (1949). *Skand. AktuarTidskr.* **33**, 119.
Hojo, T. (1931). *Biometrika*, **23**, 315.
Pearson, K. (1920). *Biometrika*, **13**, 113.

A note on Wilcoxon's and allied tests

By F. N. DAVID
University College, London

1. It is assumed that there are available two samples n_1 and n_2 randomly and independently drawn from each of two populations supposedly identical under the null hypothesis. The n_1+n_2 joint sample values are ranked in ascending order of magnitude thus forming a random sequence of two alternatives in which order must be taken into account. Mann & Whitney's adaptation of Wilcoxon's test consists, in essence, of taking the sum of the ranks of one of the alternatives. This criterion is used to test for the equivalence of the parameters of location of the two populations. Rosenbaum (1953) and Kamat (1956) have also discussed criteria based on the random sequence which may be used, if desired, to investigate

the equivalence of the parameters of dispersion of the two populations. It is the purpose of this present note to suggest a method of approach to tests on the random sequence whereby symmetric functions of the ranked observations may be used. Since such an approach is based on approximating to the distribution of a criterion by means of its moments, it will only be useful for sequences of moderate length and upwards.

2. The random sequence of length $n_1+n_2=N$, in which we take account of order, can be considered as a finite population of the first N integers. The ranks of the first sample will be n_1 elements randomly drawn from this finite population without replacement. The moments of any symmetric function of the first sample can therefore be written down immediately either from Isserlis's results (1931) or by using the procedure devised by Irwin & Kendall (1944) or from Wishart's tables (1952). Such moments will be in terms of the K-statistics of the first N integers which can themselves be obtained from the moments of the first N integers. Thus

$$K_{2r+1}=0 \quad (r=1,2,\ldots).$$

$$K_1=\frac{N+1}{2}, \quad K_2=\frac{N(N+1)}{12}, \quad K_4=-\frac{N^2(N+1)^2}{120},$$

$$K_6=\frac{N^3(N+1)^3}{252}-\frac{N^2(N+1)^2}{504},$$

$$K_8=-\frac{N^2(N+1)^2}{720}\{3N^2(N+1)^2-4N(N+1)+2\},$$

and so on.

3. *The mean of the ranks of one sample.* Abdel-Aty (1954) has shown that if we write

$$M(1^r)=\mathscr{E}_N(k_1-K_1)^r,$$

$$A_r=\alpha^{r-1}-\frac{\alpha^{r-2}}{N}+\frac{\alpha^{r-3}}{N^2}-\ldots-(-1)^{r-2}\frac{\alpha}{N^{r-2}},$$

$$\alpha=\frac{1}{n}-\frac{1}{N},$$

where N is the number in the finite population and n that in the sample, then

$$M(1)=0, \quad M(1^2)=K_2A_2, \quad M(1^3)=K_3A_3, \quad M(1^4)=K_4A_4+3K_{22}A_2^2.$$

Let $x_1, x_2, \ldots, x_{n_1}$, be the ranks of the individuals of the sample of n_1, and $\bar{x}_1$ the mean of these ranks. Then we have

$$\mu_1'(\bar{x}_1)=\frac{N+1}{2}, \quad \mu_2(\bar{x}_1)=\frac{n_2(N+1)}{12n_1}, \quad \mu_3(\bar{x}_1)=0,$$

$$K_4(\bar{x}_1)=-\frac{(N+1)n_2}{120n_1^2}\left[\frac{Nn_2}{n_1^2}+1-\frac{2N}{n_1}\right].$$

The distribution of $\bar{x}_1$ will tend to normality as N increases provided n_1/N remains fixed. To test therefore for equivalence of the population parameters of location with the alternative hypothesis that they may be different we would take the test criterion

$$\frac{\bar{x}_1-\left(\frac{N+1}{2}\right)}{\left(\frac{n_2(N+1)}{12n_1}\right)^{\frac{1}{2}}},$$

and refer it to normal tables. This is Mann & Whitney's procedure for N reasonably large and neither n_1 nor n_2 small.

4. The variance of a sample of n_1 individuals drawn from a finite population of N we define as

$$k_2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x}_1)^2.$$

The moments of k_2 have already been given in the form of multiple subscript K's by Wishart. Thus

$$M(2)=0,$$

$$M(2^2)=\frac{K_4}{n_1}+2\frac{K_{22}}{n_1-1}+(K_{22}-K_2^2),$$

with longer expressions for $M(2^3)$ and $M(2^4)$ which we do not reproduce here. Irwin & Kendall give the second moment as

$$M(2^2) = \mathscr{E}_N(k_2 - K_2)^2 = \frac{n_2(Nn_1 - N - n_1 - 1)}{n_1(n_1-1)N(N+1)} K_4 + \frac{2n_2}{(n_1-1)(N+1)} K_2^2.$$

To test whether there is a difference in dispersion one criterion could be to take the ratio k_2/K_2 or, since K_2 is a constant, just k_2. The momental constants of the ratio for sequences of 20 and 30 are given in Table 1. The standardized percentage points are obtained by interpolation from Table 42 of Pearson & Hartley (1954). For n_1 and n_2 very different it is clear that an assumption of normality for the distribution of a sum of squares will be inadequate unless the sequence is of length at least greater than 30.

Table 1. *Momental constants of k_2/K_2 for various values of N and n_1*

N	n_1	$M(2^2)/K_2^2$	$(M(2^2)/K_2^2)^{\frac{1}{2}}$	$\sqrt{\beta_1}$	β_1	β_2	$2\frac{1}{2}$ % points	
							Upper	Lower
20	10	0·0522	0·2284	0·023	0·0005$^+$	2·746	1·95	−1·93
	12	0·0334	0·1828	−0·047	0·002	2·749	1·92	−1·96
	14	0·0209	0·1445	−0·122	0·015$^-$	2·730	1·87	−2·00
	16	0·0119	0·1092	−0·219	0·048	2·674	1·79	−2·04
30	15	0·0319	0·1786	0·008	0·000	2·843	1·95	−1·94
	18	0·0207	0·1439	−0·045	0·002	2·842	1·92	−1·97
	21	0·0131	0·1142	−0·103	0·011	2·828	1·89	−2·00
	24	0·0075$^+$	0·0866	−0·180	0·032	2·788	1·885	−2·03

Mood (1954) suggests as a rank criterion to test dispersion, the sum of squares of the deviations of one sample from the population mean, $\frac{1}{2}(N+1)$. This is

$$\sum_{i=1}^{n_1} (x_i - K_1)^2 = (n_1 - 1)k_2 + n_1(k_1 - K_1)^2,$$

the moments of which can be written down immediately from Wishart's tables. Which criterion is the more sensitive to changes in the dispersions of the two populations generating the samples it does not appear as yet possible to say. It is also difficult to see whether a criterion based on the sums of squares of both samples, e.g.

$$k_2(1) + k_2(2)$$

will be an improvement on a criterion based on a single sample, but probably little sensitivity will be gained if this is tried. Whatever criterion is used it will be possible from Wishart's tables to write down its moments either exactly or to any order of approximation desired and to approximate to its probability distribution (for sequences of reasonable length) by means of a smooth curve.

Tests based on symmetric functions of all the ranks of a single sample should be more discriminating than tests based on ordered variables, such as range for example. One point should, however, be mentioned about all ranked tests for dispersion differences. It may be that at the same time as there is a change in the dispersion parameters of the populations generating the sample, there is also a change in the location parameters which will effectively mask it. It does not seem possible to devise a single criterion based on ranks which would take this into account.

5. *Correction for ties.* Corrections for ties have been discussed recently by Putter (1955) who gives a list of references. Provided the sequence is long enough and the ties not too frequent a straightforward and easy allowance can be made. Suppose we have v elements which tie at rank r. The usual method of assigning ranks is to give each of the v elements the same rank, $r + \frac{1}{2}(v-1)$. K_1 is then unaltered but all the other K-statistics of the population are changed. We can calculate the K-statistics of this changed population and use these values in the moments of whatever criterion we are interested in. For n_1 and n_2 large and for the number of sets of ties not too small the effect appears negligible.

REFERENCES

ABDEL-ATY, S. H. (1954). *Biometrika*, **41**, 253.
IRWIN, J. O. & KENDALL, M. G. (1944). *Ann. Eugen., Lond.*, **12**, 138.
ISSERLIS, L. (1931). *Proc. Roy. Soc.* A, **132**, 586.
KAMAT, A. R. (1956). *Biometrika*, (in the Press).
MOOD, A. M. (1954). *Ann. Math. Statist.* **25**, 514.
PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
PUTTER, J. (1955). *Ann. Math. Statist.* **26**, 268.
ROSENBAUM, S. (1953). *Ann. Math. Statist.* **24**, 663.
WISHART, J. (1952). *Biometrika*, **39**, 1.

Likelihood function for capture-recapture samples

BY N. E. G. GILBERT
John Innes Horticultural Institution

When capture-recapture methods are used on a wild population, the sampling fraction is usually small, so that the distribution of marked and unmarked individuals in the recapture samples can be treated as binomial. However, samples of insects have sometimes included the greater part of the population. When a chain of such samples is used to estimate several statistics (e.g. population size and birth- and death-rates), the appropriate hypergeometric distribution is intractable. This note describes a simple way of correcting the error which is introduced when the binomial distribution is used instead of the hypergeometric. The population of size N contains n marked individuals, and one sample in the chain is of size X of which x are marked. The situation will, of course, be more complicated than this in practice.

Chapman (1951) suggests minimizing

$$\chi^2 = \Sigma \frac{X(x/X - n/N)^2}{(x/X)(1-x/X)(1-x/n)},$$

where the summation extends over all samples in the chain. These estimates are asymptotically efficient; but since x/X is used instead of n/N in the denominator, they can be improved upon in the present case where information about N is obtainable from the other samples in the chain. This point is illustrated in the numerical example.

A better weighting of the contributions from the different samples is obtained by minimizing the more usual expression for χ^2 with denominator $(n/N)(1-n/N)(1-X/N)$, but the solution of the resulting equation involves heavy arithmetic.

Using the binomial approximation, the likelihood function is

$$\log L = \Sigma\{x \log(n/N) + (X-x)\log(1-n/N)\},$$

so that

$$\frac{d(\log L)}{dN} = \Sigma \frac{nX - Nx}{N(N-n)}.$$

This expression may also be obtained by differentiating the numerator only of

$$\chi^2 = \Sigma \frac{(Nx - nX)^2}{Nx(N-n)} \quad \text{w.r.t. } N.$$

In the hypergeometric case, each term in this expression for χ^2 is divided by a further $(1-X/N)$. The maximum-likelihood equation, when similarly modified, becomes

$$\frac{d(\log L)}{dN} = \Sigma \frac{nX - Nx}{(N-X)(N-n)},$$

and the resulting estimates differ negligibly from those obtained by using the hypergeometric likelihood, although the arithmetic involved is obviously no worse than for the binomial case.

Example. Suppose that eight individuals from a population with zero births and deaths are marked and released; that a first recapture sample of eight contains five marked individuals, so that the total number marked is now eleven; and that a second sample of thirteen contains five marked individuals. (With zero births and deaths, the distribution of the two kinds of mark affords no extra information.)

This provides a simple numerical illustration; the two sampling fractions are different but not too much so (i.e. the estimation is not dominated by the information from one sample), and the two estimates of N from the individual samples differ (but not significantly).

The estimates are

Chapman's minimum χ^2	$\hat{N} = 20{\cdot}4$
Binomial	$\hat{N} = 21{\cdot}7 \pm 4{\cdot}91$
Modified binomial	$\hat{N} = 23{\cdot}1 \pm 3{\cdot}68$
Hypergeometric	$\hat{N} = 23{\cdot}2 \pm 3{\cdot}54$

REFERENCE

CHAPMAN, D. G. (1951). *Univ. Calif. Publ. Statist.* **1**, 131.

Tables of Poisson power moments

BY J. B. DOUGLAS

School of Mathematics, N.S.W. University of Technology, P.O. Box 1, *Kensington, New South Wales*

In the fitting of the complete and truncated Neyman Type A contagious distribution by maximum-likelihood methods, the rather tedious arithmetic can be considerably reduced by use of tables of moments about the origin of the Poisson distribution. Appropriate tables, to 3 or 4 (and occasionally 5) significant digits, with a description of their method of use, were published in *Biometrics*, **11** (1955), 149–73.

The tables are in fact of ratios of moments. If

$$\mu'_x = e^{-\lambda} \sum_{r=0}^{\infty} r^x \lambda^r / r!,$$

then

$$p_x = \mu'_{x+1}/\mu'_x$$

and

$$q_x = p_x(p_{x+1} - p_x)$$

are tabulated for

$$x = 0\,(1)\,19,$$

$$\lambda = 0{\cdot}000\,(0{\cdot}001)\,0{\cdot}03\,(0{\cdot}01)\,0{\cdot}3\,(0{\cdot}1)\,3{\cdot}0.$$

However, these tables also exist in manuscript form with values of p_x calculated correct to 6 or 7 significant digits (and corresponding values of q_x), the ranges of tabulation being those given above. The existence of these more detailed tables is therefore recorded here; the author would be glad to learn of any applications where their use would be desirable.

REVIEWS

Theory of Games and Statistical Decisions. By D. BLACKWELL and M. A. GIRSHICK.* New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1954. Pp. xi + 355. 60*s*.

Readers of this journal will have seen Barnard's detailed review of A. Wald's *Statistical Decision Functions* (*Biometrika*, 1953, **40**, 475–7). The present book shows how statistical problems of various types can be handled via games theory and decision functions. It is intended as a text for first-year graduate students who may, however, find it heavy going; they will be familiar with the statistics required, but not with a number of other topics mostly of a topological nature. The authors have tried to make the book self-contained as regards most of these, but even so it will never be light reading. The detailed bibliography provided, together with a reading list for each chapter, will help to overcome some of these difficulties.

The first few chapters are devoted to a thorough exposition of games theory; since the whole treatment is in terms of that theory this was necessary. Even so, a preliminary reading of McKinsey's *Introduction to the Theory of Games* will be useful. A general description of statistical games (i.e. of statistical problems viewed as games between nature and the statistician) rounds off this part. This is followed by a chapter on utility and principles of choice which is, unfortunately, too short to provide more than a statement of utility theory in Ramsey–von Neumann terms and of the principles of choice known as minimax, minimax regret and Bayes. The reader is not made aware that other 'reasonable' criteria exist as well—presumably the authors had to draw a line somewhere. The rest of the book deals with theory of estimation, built on those foundations. There are two chapters on fixed sample-size games and two on sequential procedures. The final chapters are concerned with theory of estimation and comparison of experiments; they are most interesting and rewarding.

The reader will be pleasantly surprised to find that his way of making decisions following classical theory is well founded even by the exacting standards set by the authors. It should be noted that 'decisions' include quite properly estimation and not only design of experiments or sequential procedures. The fact that decision theory supports normal statistical habits is reassuring—which does not mean that it is an unnecessary elaboration of the obvious, since a number of new applications follow from the new outlook.

The practical man will probably require another book giving more detailed recipes and leaving out some of the rigours. Ultimately he will have to be grateful to the authors whose work makes the writing of such a book possible. The quest for generality has made the present book more difficult to read than it should be (see, for example, theorem 2.7.2 and its proof), but this might be considered a merit in an advanced text. If decision theory resembles a steam hammer it is also true that the statistician's universe is not all nuts.

The book contains many stimulating exercises and is well produced. G. MORTON

Elementary Statistics for Students of Social Science and Business. By R. C. SPROWLS. London: McGraw-Hill Book Co. Inc. 1955. Pp. 392. 41*s*. 6*d*.

Basic Statistical Concepts. By J. K. ADAMS. London: McGraw-Hill Book Co. Inc. 1955. Pp. 304. 41*s*. 6*d*.

Elementary Statistics for Students of Social Science and Business is a text-book on the rudiments of statistics and probability with most of the mathematical techniques omitted. For this reason much explanation and example are necessary and these the author gives. Nine chapters are devoted to probability (including the binomial distribution) and methods of measuring location and dispersion, and the tenth covers linear regression. Thus far the statistics will be elementary for all. The author, who is an Assistant Professor of Business Statistics, then shows his bias by discussing the analysis of

* It is sad to note that M. A. Girshick died on March 2, 1955 at the early age of 46, with his research and teaching activities in full swing. An obituary appeared in the June 1955 number of *The American Statistician*.

time series and index numbers, two branches of statistics favoured by economists and business men. At the level at which the author aims the book is reasonably adequate. It is frankly utilitarian and conveys nothing of the excitement and interest attendant on the analysis of statistical problems. Certain topics might well have been omitted or at least accurately discussed. For example, it is questionable whether the statement that maximum-likelihood estimates are unbiased is sensible, since often maximum likelihood and bias seem to go hand in hand.

The reviewer would judge this book as covering much of the work which would be done by a first year student in an English University reading statistics ancillary to another subject. Such students undoubtedly find it easier to think in words rather than symbols and by them this book may be found helpful.

Mr Adams, who writes *Basic Statistical Concepts*, is an Assistant Professor of Psychology. This book is not, however, a combination of factor analysis and 'stat-rats' as might possibly be expected, but a straightforward exposition of elementary statistics covering, from a less mathematical point of view, much the same ground already covered by A. M. Mood, *Introduction to the Theory of Statistics* (McGraw-Hill). There is no question in the reviewer's mind which book is to be preferred. Given the writer's objective, however, the exposition is tolerable if uninspired, and may be suitable as a textbook for the type of student which he has in mind. It is interesting to note that in our parent subject of mathematics every teacher of mathematics does not find it necessary to write a text-book for the use of his classes—in fact very few do. Immediately after the war when the spate of statistical text-books began it seemed to be a tribute to the virility of our subject that so many persons should feel the urge of exposition. Looking back over the decade one realizes that it is probably because the interpretation of statistical mathematics is essentially subjective, and to each teacher the techniques and basic framework mean something slightly different. Nevertheless, the reviewer would suggest that the time has come to call a halt. Enough books on statistical methods now exist to suit most points of view even if it is necessary to indicate while teaching different emphasis where it appears desirable. In ten years we have grown up and what our subject now requires is not more and more verbiage on the elements but a series of specialist monographs each on a separate branch, a process which is already under way in the design of experiments.

F. N. DAVID

Experimental Design and its Statistical Basis. By D. J. FINNEY. London: Cambridge University Press. 1955. Pp. xi+169. 30*s*.

Experimental Design. Theory and Application. By W. T. FEDERER. New York and London: The Macmillan Company. 1956. Pp. 544+47. 77*s*.

It is said that Beethoven when young divided his time between the study of counterpoint and the composition of music without becoming aware that the one was supposed to be an aid to the other. Many applied statisticians will sympathize with him in this. They spend their lives trying to understand the purpose and background of research programmes, debating the merits of different sets of treatments, and assessing the need to make a certain comparison especially accurate, and usually conclude by designing something quite simple in randomized blocks. In books, however, they live in a world of hyper-graceo-latin-cubes and super-magic latin squares; while, if Prof. Federer's early pages are to be believed, the selection of treatments and of characteristics to be measured are 'non-statistical'. He may be right, but Dr Finney would not agree.

Dr Finney's book is at once remarkable and remarkably good. It aims not at teaching statistics but at explaining what statisticians do. It discusses experimental design from the point of view of logic and common sense, and shows how many scientific arguments can be expressed in numerical form. There would be little point in a chapter-to-chapter critique; it will perhaps suffice to say that there is now at least one statistical laboratory where biologists asking for instruction will be asked first to read this book. They will not understand the chapter on fractional replication and confounding, but they should thereafter see statistical methods as a systematization of what they have previously done unmethodically, not as an alien culture imposed by a conquering power on the ancient civilization of medical and biological science.

Prof. Federer's book is different in purpose; it deals with experimental design only in the narrowest methodological sense, and within those limits is concerned to be comprehensive and detailed, which it certainly is. Also, it has a bibliography that is well-chosen and extensive. Nevertheless, for all its merits this is a disconcerting book to read. Thus, on p. 307 there starts a passage about 'lattices' and 'incomplete blocks' that the reviewer still found incomprehensible after much rereading, nor was he

helped by finding a table subjoined headed, 'Incomplete blocks in a randomized complete block design'. Light dawned on p. 314 from a passage, which in any other context would have been a masterpiece of obscurity, from which it was gathered that the two terms were synonymous, and that incomplete blocks in complete block designs were what most people call lattices. This is perhaps an extreme example, but others nearly as bad could be adduced. Nevertheless, on account of its detailed completeness this book is to be commended. S. C. PEARCE

Population Genetics. By CHING CHUN LI. London: Cambridge University Press, for University of Chicago Press. 1955. Pp. xi+366. 75*s*.

This book is a summary of the mathematical theory of population genetics. It begins with the study of large random-mating populations, and the Hardy–Weinberg law. It goes on to the consideration of correlations between relatives, and the modifications required for linked and sex-linked genes, and for autopolyploids. Various schemes of inbreeding are then discussed, such as selfing and sib-mating. Path coefficients are explained, and used in the study of correlation and more complicated systems of inbreeding. The author goes on to discuss the effects of assortative mating, mutation, selection, subdivision and migration. Finally, there are two chapters on the random fluctuations which are inevitable in small populations. In short, most of the subjects of major importance in human and population genetics are included, although in some cases the treatment is disappointingly brief, presumably in order to keep the size of the book within reasonable limits. But it would be very interesting to have more information on, for example, polyploid segregations. The emphasis is entirely on the mathematical theory, and only a little is said about the practical analysis of actual data.

The explanations are extremely lucid, and anyone with a knowledge of quite elementary statistics and genetics should be able to follow most of the book without difficulty (even without calculus). The reviewer has noticed very few errors, although the proof of the 'simplified version' of Fisher's fundamental theorem on natural selection (p. 273) is incorrect as it stands, since Li forgets to include the change in fitness between one generation (after selection) and the next generation (at birth). Also the final discussion on pp. 344–7 is a little confused, and may not take fully into account the complications introduced by inbreeding, linkage and assortative mating. And Li does not mention that the exact form of at least some of the gene-frequency distributions is rather controversial. But these are rather minor blemishes, and on the whole Li has set out the argument very clearly and logically, and is also well aware that the mathematical models represent an oversimplified version of the true biological situation.

The printing is excellent, and the diagrams are clear and helpful. There is a useful bibliography at the end. This book can be warmly recommended to anyone interested in the mathematical treatment of the genetical structure and evolution of populations. CEDRIC A. B. SMITH

Statistical Methods (third edition). By FREDERICK C. MILLS. London: Sir Isaac Pitman and Sons, Ltd. 1955. Pp. xviii+842. 50*s*.

This third edition of Prof. Mills's well-known text-book has been substantially rewritten 'to take account of the more important of recent developments (in statistical theory) that bear on the applications of statistics in the social sciences, in business administration, and in governmental affairs'. It is addressed to the non-mathematical reader. Concepts and formulae are stated, and their applications explained, with examples (based, as is to be expected in an American text-book, on American statistics) worked in detail; the mathematical reasoning on which the formulae are based is generally omitted. Among the topics included are the description of frequency distributions, problems of estimation, tests of hypotheses, the analysis of variance, simple and multiple correlation, methods of sampling, the analysis of time series, and index-numbers.

The presentation is clear and unambiguous without being unusually inspiring, and the book should be intelligible to any reasonably intelligent reader who has determined to understand the elements of statistical method; but, as in many other books on the subject, the demand on the reader's acquaintance with elementary algebra varies from chapter to chapter. Inevitably in an exposition of this scope, the treatment of relatively advanced topics omits some important qualifications and special cases; and in this, of course, lies the danger that the student may misapply the methods.

This is not a book—nor is it claimed to be one—for the industrial statistician; quality control and production control are not treated, nor is the design of experiments. More surprisingly, since the book

is intended to deal with the use of statistics in business administration, there is no reference to market research; and while random, stratified, systematic and cluster sampling are described, I have found no reference, even in condemnation, to quota sampling. One might also have expected, in a work at this level, some discussion of the collection of statistical data, the importance of definitions, margins of observational error, and similar matters which it is essential to consider when dealing with social, economic, business and administrative statistics.

The perfect text-book of elementary statistics still remains to be written, and perhaps it is an impossibility, but the book under review will suit very many students. Their attention is directed to the 'errata sheet' which precedes the first page of the text; this relates to the biblography on pp. 821–30, from which, by an oversight, the numbers were omitted without which references in the text are meaningless. FREDERICK BROWN

The Art of Investment. By A. G. ELLINGER. London: Bowes and Bowes. 1955. Pp. 170. 15*s*.

Part I of Mr Ellinger's book briefly sketches the mechanisms of the London Stock Markets, and details the recent history of security price-movements. Parts II and III are concerned with investment strategies and decisions, considered as functions of the movements in charts of prices, yields and volumes of share trading. These charts are taken as embodying the collective wisdom of investors. Since it is unprofitable to invest against market trends, the practical problem is to set up criteria for deciding the directions of the main trends. Mr Ellinger's system is simple, although hedged about with rather imprecise qualifications and exceptions. It depends ultimately on the linearity of trends in the time series, charted on a logarithmic scale. (The unattractive consequence, that the trends on a natural scale are exponential, is not discussed.)

The important question is: does Mr Ellinger's system work reasonably well? The evidence presented is suggestive, but far from conclusive. I should like Mr Ellinger to formulate his system categorically and exhaustively, and apply it publicly to a random sample of securities selected by someone else. The results would be more convincing than any amount of examination of the past. A. STUART

Measuring Business Changes; a Handbook of Significant Business Indicators. By RICHARD M. SNYDER. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1955. Pp. 382. 64*s*.

Mr Snyder's reference book describes and explains about fifty available economic 'indicators' for the United States economy, ranging over National Income, Population Growth, Labour Statistics, Commodity Prices, Indices of Production and Activity (both general and specific to some industries), Domestic and International Trade, Finance and Stock Markets. It is clearly an indispensable source for anyone seeking published information in this field, and gives ground for hope that a similar single source book will become available for the United Kingdom. A. STUART

Statistics (second edition). By L. H. C. TIPPETT. Oxford University Press. 1956. Pp. 224. 7*s*. 6*d*.

This book, already a classic in the 'expositions-for-the-layman' group, has been revised and brought up to date by the author. The general plan and coverage of the book remain unchanged.

F. N. DAVID

Tafeln zum Vergleich zweier Stichproben mittels X-Test und Zeichentest (Tables for comparing two samples by X-Test and Sign Test.) By B. L. VAN DER WAERDEN and E. NIEVERGELT. Berlin: Springer-Verlag. 1956. Pp. 34. DM. 4.80.

This booklet describes the method of computation and reference to tables of the pseudo-'t' order-statistic 'X', which van der Waerden proposed as a test of difference of central tendency in two samples (*Math. Ann.* (1953), **126**, 93). A similar discussion of the sign test analogue of the matched-pairs t-test is also given.

Part II (in German) gives a worked example and Part IV is the English translation of this. Part III consists of tables (largely derived more or less directly from standard normal and binomial tables) and Part I gives, in German, a discussion of the relative merits of different tests and the method of computation of the tables. Since the X-test is applicable in exactly the same circumstances as Mann & Whitney's generalization of Wilcoxon's test and since, as van der Waerden notes in his comparative paper (*Proc. K. Acad. Wet. Amst.* A (1953), **56**, 310), the asymptotic relative efficiency of this as compared with Student's t is $\sqrt{(3/\pi)} = 0{\cdot}977$ when there is normal variation, the reviewer at least was not convinced of the need for a new test nor of its superiority to Wilcoxon's.

D. E. BARTON

PUBLICATIONS OF THE U.S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS

(i) Tables of the error function and its derivative. Applied Mathematics Series 41. 1954. [Re-issue of the 1941 Mathematical Table No. 8.] Pp. xi + 302. $3.25.

The main table provides values of the functions

$$H(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\alpha^2}d\alpha \quad \text{and} \quad H'(x) = \frac{2}{\sqrt{\pi}}e^{-x^2}$$

to 15 decimal places, for argument $x = 0\,(0{\cdot}0001)\,1{\cdot}0000\,(0{\cdot}001)\,5{\cdot}600$. A short supplementary table gives values of $1 - H(x)$ and $H'(x)$ to eight significant figures for $x = 4{\cdot}00\,(0{\cdot}01)\,10{\cdot}00$.

(ii) Table of sine and cosine integrals for arguments from 10 to 100. Applied Mathematics Series 32. 1954. [Re-issue of the 1942 Mathematical Table No. 13.] Pp. xv + 187. $2.25.

The main table provides values of the functions

$$\mathrm{Si}\,(x) = \int_0^x \frac{\sin t}{t}dt \quad \text{and} \quad \mathrm{Ci}\,(x) = \int_\infty^x \frac{\cos t}{t}dt$$

with second differences, to 10 decimal places, for argument $x = 10{\cdot}00\,(0{\cdot}01)\,100{\cdot}00$. In addition, to facilitate interpolation, auxiliary tables are given of $\frac{1}{2}p(1-p)$ and $\frac{1}{6}p(1-p^2)$ for $p = 0\,(0{\cdot}001)\,1{\cdot}000$. There is also a table to 15 decimal places of multiples of $\frac{1}{2}\pi$, i.e. of $\frac{1}{2}n\pi$ for $n = 1\,(1)\,100$.

(iii) Table of the gamma function for complex arguments. Applied Mathematics Series 34. 1954. Pp. xvi + 105. $2.00.

This table gives the real and imaginary parts of $\log_e \Gamma(z)$ for $z = x + iy$, and $x, y = 0{\cdot}0\,(0{\cdot}1)\,10$, each to 12 decimal places. Auxiliary tables of $\sin \pi x$, $\cos \pi x$, $\sinh \pi x$ and $\cosh \pi x$ are given to 15 decimal places (or 15 significant figures) for $x = 0\,(0{\cdot}1)\,10{\cdot}0$ to facilitate extension of the scope of the table.

OTHER BOOKS RECEIVED

Famous Problems and other Monographs. By F. KLEIN *et al.* New York: Chelsea Publishing Company. 1955. Pp. 321. $3.25.

Income of American People. By H. P. MILLER. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1955. Pp. xvi + 206. 44*s*.

Analysis of Confounded Factorial Experiments in Single Replication. By F. E. BINET *et al.* North Carolina Agricultural Experiment Station (Technical Bulletin No. 113). 1955. Pp. 64.

Theoretical Genetics. By R. B. GOLDSCHMIDT. London: Cambridge University Press, for University of California Press. 1956. Pp. 563. 64*s*.

Health in Industry (Sickness Absence Statistics). For London Transport Executive by Butterworth's Medical Publications. 1956. Pp. 177. 35*s*.

Elementary Topology. By D. W. HALL and G. L. SPENCER. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1955. Pp. 303. 56*s*.

Rectangular-Polar Conversion Tables. By E. H. NEVILLE. London: Cambridge University Press, for the Royal Society. 1956. Pp. xxxii + 109. 30*s*.

Irrationalzahlen, 2nd edition. By O. PERRON. New York: Chelsea Publishing Co. 1951. Pp. 199. $1.50.

Trigonometrical Series, 2nd edition. By A. ZYGMUND. New York: Chelsea Publishing Co. 1952. Pp. 329. $1.50.

Statistics. By W. A. WALLIS and H. V. ROBERTS. Glencoe, Illinois: The Free Press. 1956. Pp. xii + 646. $6.

The Essentials of Educational Statistics. By FRANCIS G. CORNELL. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1956. Pp. 375. 46*s*.

Annual Epidemiological and Vital Statistics 1953. World Health Organization, Geneva, Switzerland. (U.K. Sales Agent, H.M.S.O.) 1956. Pp. 571. 50*s*., $10 or Sw. fr. 30.

CORRIGENDA

Correlated Random Normal Deviates. Tracts for Computers, No. 26. By E. C. FIELLER, T. LEWIS and E. S. PEARSON.

These tables were issued by the Department of Statistics, University College, London, and published by the Cambridge University Press in 1955. As an experiment in method of reproduction, the tables were not set up in type but were reproduced direct by photo-offset process from sheets prepared at the National Physical Laboratory on an electromatic typewriter. It was hoped that this method would provide a clear facsimile copy of the original sheets, but it has been found unfortunately that this is not the case and a number of errors have appeared in the tables.* These consist of

(*a*) missing or broken negative signs,
(*b*) broken figures,
(*c*) one incorrect figure and one addition of a negative sign, probably due to printers' 'touching up'.

These errors or opportunities for the misreading of figures are more serious in the present table than they would have been in some tables of random numbers or random normal deviates, because on each page column totals have been included in the form of Σx, Σx^2, $\Sigma(x_0 x_t)$ and $r(x_0, x_t)$. Thus there is a risk of inconsistent results arising from certain uses of the figures.

The 60 pages of published table have now been carefully checked with the original sheets. The errors found falling under headings (*a*) and (*c*) are listed below in full; for (*b*) only those cases in which real ambiguity seems likely to arise are given.

A full correction will be made to the figures in the second impression of the Tract which will be issued during the coming year. In the meantime, the authors and publishers wish to express their sincere regret to all those who have purchased or used the Tract in its present form. E.C.F., T.L., E.S.P.

(*a*) *For the values listed here the negative sign is completely missing, or less than half there*

Page	Column	Row	Page	Column	Row
1	x_1	24	22	x_2	30
1	x_8	6*	23	x_1	34
3	x_5	19*	24	x_1	46
4	x_3	25	28	x_3	1
4	x_4	47 and 48	30	x_8	44
5	x_9	46	32	x_1	10
6	x_2	47	32	x_4	5
6	x_7	19	35	x_4	44*
7	x_0	50	36	x_0	22
8	x_6	42*	37	x_6	29
11	x_7	47*	37	x_7	24
11	x_9	15	40	x_7	37
13	x_0	33*	44	x_1	36*
13	x_2	24	47	x_5	43*
13	x_2	25	52	x_8	44
13	x_4	46	53	x_4	43
13	x_5	19	54	x_2	15*
16	x_3	10	56	x_0	19
18	x_9	37	56	x_4	8
19	x_0	45*	56	x_5	16
19	x_1	25	56	x_5	31
19	x_1	32	56	x_9	42
21	x_1	10	57	x_3	46
21	x_6	41	59	x_4	15
21	x_9	47	59	x_4	26

* Negative sign missing altogether.

* The process of reproduction did not easily allow the customary checking in proof which occurs with a printed table.

(*b*) *These values all have broken figures which might be misread*

Page	Column	Row	Should read
5	x_7	$\Sigma x_0 x_7$	49·8068
12	x_9	40	−0·37
16	x_9	Σx	−5·95
16	x_4	19	0·83
25	x_5	Σx	−6·54
28	x_2	7	−1·53
34	x_7	40	−1·36
36	x_2	10	2·20
41	x_3	44	delete mark in front of number which might be taken as a minus sign. Read 1·01
48	x_7	1	0·66
57	x_1	50	0·37
60	x_4	9	1·88

(*c*) *Incorrect additions*

6	x_0	43	For −0·60 read 0·60
28	x_2	19	For −0·53 read −0·54

J. GANI. 'Some theorems and sufficiency conditions for the maximum likelihood estimator of an unknown parameter in a simple Markov chain.' *Biometrika*, **42**, 342–59.

The statement in line 8 below equation (22) on p. 352 is wrong, and invalidates the proof leading to (23); this result can be obtained more simply as follows.

Instead of considering the probabilities $p(x_i, \theta)$ that the ith observation of the variate x be x_i, write $p_j(\theta)$ as the probability that the variate x has the value j, where j is some integer in the given range. Let n_j be the frequency with which the value j occurs in the sample, so that $\sum_j n_j = n$, the total number of observations. Writing the likelihood function as

$$L = \sum_j n_j \ln p_j(\theta),$$

the factorizability condition gives

$$\sum_j n_j \ln p_j(\theta) = \ln f(n_1, n_2, \ldots) + \ln F(T, \theta),$$

where $T = T(n_1, n_2, \ldots)$ is a sufficient estimator of θ. Providing $p_j(\theta)$ and $F(T, \theta)$ are differentiable with respect to θ, this differentiation will lead to

$$\sum_j n_j u_j(\theta) = \sum_j n_j \frac{d}{d\theta}\{\ln p_j(\theta)\} = \frac{\partial}{\partial \theta}\{\ln F(T, \theta)\} = G(T, \theta),$$

where $G(T, \theta)$ is some linear function of the frequencies n_j.

Now let us assume that T is a function formally differentiable in each of the n_j, and that $G(T, \theta)$ is similarly differentiable in T; differentiation of $G(T, \theta)$ with respect to n_j leads to

$$\frac{\partial G(T, \theta)}{\partial T}\frac{\partial T}{\partial n_j} = u_j(\theta) \quad (j = 1, 2, \ldots).$$

Since T is a function of the n_j only, this relation shows that $\partial T/\partial n_j$ can only be some function $1/v_j(T)$ of T, so that

$$\frac{\partial G}{\partial T} = u_j(\theta)\, v_j(T) \quad (j = 1, 2, \ldots).$$

The equality of these products for all j is possible only if

$$u_j(\theta) = c_j K_1(\theta), \quad v_j(T) = c_j^{-1} v(T),$$

where the c_j are constants, and the form of the function $G(T, \theta)$ is, therefore,

$$G(T, \theta) = K_1(\theta) \int v(T)\, dT + K_2(\theta) = K_1(\theta)\, g(T) + K_2(\theta),$$

as previously obtained in the paper. The function $g(T)$ is clearly

$$g(T) = \int v(T)\, dT = \sum_j \int v(T) \frac{\partial T}{\partial n_j} dn_j = \sum_j c_j n_j,$$

a linear function of the n_j.

The result (23), where the likelihood function is expressed in terms of individual observations x_i instead of the frequencies n_j, follows directly.

I am greatly indebted to Mr E. Bowen for pointing out my error, and to Prof. G. A. Barnard for suggesting the proof above.

JOHN WISHART

1898–1956

John Wishart, who has been connected with this journal in an editorial capacity since 1937, died suddenly in Mexico on 14 July last. An obituary and appreciation will appear in the next number.

E. S. P.

TRACTS FOR COMPUTERS

Department of Statistics, University College, London

I. Tables of the Digamma and Trigamma Functions. By ELEANOR PAIRMAN, M.A. Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i + q_1)(p_2 i + q_2) \dots (p_n i + q_n)}$ where the p's and q's are numerical factors. *Price 5s. net.*

V. Table of Coefficients of Everett's Central-Difference Interpolation Formula. By A. J. THOMPSON, PH.D. *Second edition. Price 7s. 6d. net.*

VIII. Table of the Logarithms of the Complete Γ-Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2). By EGON S. PEARSON, D.SC. *Price 5s. net.*

IX. Log Γ (x) from $x = 1$ to 50·9 by intervals of 0·01. By JOHN BROWNLEE, M.D., D.SC. *Price 5s. net.*

X. On Quadrature and Cubature or on Methods of Determining Approximately Single and Double Integrals. By J. O. IRWIN, D.SC. *Price 7s. 6d. net.*

XII. Tables of the Probable Error of the Coefficient of Correlation. By KARL HOLZINGER, PH.D. *Price 5s. net.*

XIII. Bibliotheca Tabularum Mathematicarum, being a Descriptive Catalogue of Mathematical Tables. Part I. A, Logarithms of Numbers. By JAMES HENDERSON, PH.D. *Price 9s. net.*

XV. Random Sampling Numbers. By L. H. C. TIPPETT, M.SC., with a Foreword by KARL PEARSON. *Price 5s. net.*

XXIII. Tables of $\tan^{-1}x$ and $\log(1+x^2)$. To assist in the calculation of the ordinates of a Pearson Type IV curve. By L. J. COMRIE, PH.D. *Price 5s. net.*

XXIV. Random Sampling Numbers (2nd Series). By M. G. KENDALL and B. BABINGTON SMITH. *Price 5s. net.*

XXV. Random Normal Deviates. By HERMAN WOLD. *Price 5s. net.*

XXVI. Correlated Random Normal Deviates. By E. C. FIELLER, T. LEWIS and E. S. PEARSON. *Price 10s. 6d. net.*

Nos. II, III, IV, VI and VII are out of print

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, Ph.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS'S *Arithmetica Logarithmica*).

The nine separate sections of this Table have now been issued, and the complete work consisting of the logarithms of numbers 10,000–100,000, together with Dr Thompson's General Introduction (98 pp.) is now available in two bound volumes.

Price £8. 8s. 0d.

Issued by the CAMBRIDGE UNIVERSITY PRESS, Bentley House, LONDON, N.W.1
on behalf of the
DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON
and obtainable from any bookseller

(*All rights reserved*)

BIOMETRIKA Vol. 43, Parts 3 and 4

CONTENTS

PAGE

Studies in the history of probability and statistics. III. A note on the history of the graphical presentation of data. By ERICA ROYSTON 241

Studies in the history of probability and statistics. IV. A note on an early statistical study of literary style. By C. B. WILLIAMS 248

A goodness of fit test for spectral distribution functions of stationary time series with normal residuals. By A. M. WALKER 257

Sufficiency conditions in regular Markov chains and certain random walks. By J. GANI 276

Some asymptotic distribution theory for Markov chains with a denumerable number of states. By CYRUS DERMAN 285

A general method for approximating to the distribution of likelihood ratio criteria. By D. N. LAWLEY . . 295

On the accuracy of weighted means and ratios. By G. S. JAMES 304

On estimating the latent and infectious periods of measles. II. Families with three or more susceptibles. By NORMAN T. J. BAILEY 322

Significance tests for a variable chance of infection in chain-binomial theory. By NORMAN T. J. BAILEY . 332

On the variation of yield variance with plot size. By P. WHITTLE 337

On the construction of significance tests on the circle and the sphere. By G. S. WATSON and E. J. WILLIAMS 344

Notes on bias in estimation. By M. H. QUENOUILLE 353

An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. By S. N. ROY and S. K. MITRA 361

A two-sample distribution-free test. By A. R. KAMAT 377

Sequential analysis applied to certain experimental designs in the analysis of variance. By W. D. RAY . 388

Lognormal approximation to products and quotients. By S. R. BROADBENT 404

A rejection criterion based upon the range. By C. I. BLISS, W. G. COCHRAN and J. W. TUKEY . . . 418

Confidence intervals for a proportion. By EDWIN L. CROW 423

Serial correlation in regression analysis. II. By G. S. WATSON and E. J. HANNAN 436

MISCELLANEA

Revised upper percentage points of the extreme studentized deviate from the sample mean. By H. A. DAVID 449

Exact linear sequential tests for the mean of a normal distribution. By J. TAYLOR 452

On the sum of squares of normal scores. By H. RUBEN 456

On the moments of the range and product moments of extreme order statistics in normal samples. By H. RUBEN 458

On estimating binomial response relations. By F. J. ANSCOMBE 461

Existence and uniqueness of a uniformly most powerful randomized unbiased test for the binomial. By A. A. BLANK 465

A note on the circular multivariate distribution. By G. S. WATSON 467

The fitting of regression curves with autocorrelated data. By N. A. HUTTLY 468

Bounds for the variance of Kendall's rank correlation statistic. By ALAN STUART 474

A note on the theory of quick tests. By D. R. COX 478

A note on the signs of gross correlation coefficients and partial correlation coefficients. By OLAV REIERSØL 480

The estimation of the mean of a censored normal distribution by ordered variables. By P. G. MOORE . 482

A note on Wilcoxon's and allied tests. By F. N. DAVID 485

Likelihood function for capture-recapture samples. By N. E. G. GILBERT 488

Tables of Poisson power moments. By J. B. DOUGLAS 489

REVIEWS

D. BLACKWELL and M. A. GIRSHICK's 'Theory of Games and Statistical Decisions' 490

R. C. SPROWLS's 'Elementary Statistics for Students of Social Science and Business' 490

J. K. ADAMS's 'Basic Statistical Concepts' 490

D. J. FINNEY's 'Experimental Design and its Statistical Basis' 491

W. T. FEDERER's 'Experimental Design. Theory and Application' 491

CHING CHUN LI's 'Population Genetics' 492

FREDERICK C. MILLS's 'Statistical Methods' 492

A. G. ELLINGER's 'The Art of Investment' 493

RICHARD M. SNYDER's 'Measuring Business Changes; a Handbook of Significant Business Indicators' . 493

L. H. C. TIPPETT's 'Statistics' 493

B. L. VAN DER WAERDEN and E. NIEVERGELT's 'Tafeln zum Vergleich zweier Stichproben mittels X-Test und Zeichentest' 493

NATIONAL BUREAU OF STANDARDS, Publications of the U.S. Department of Commerce, Applied Mathematics Series 41, 32, 34 494

OTHER BOOKS RECEIVED 495

CORRIGENDA 496

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

PAGE
241
248
257
276
285
295
304
322
332
337
344
353
361
377
388
404
418
423
436
449
452
456
458
461
465
467
468
474
478
480
482
485
488
489
490
490
490
491
491
492
492
493
493
493
493
494
495
496